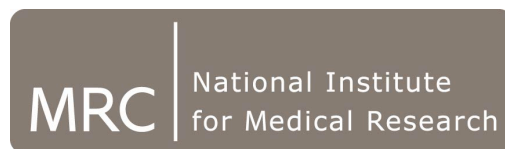


Regulation of Neural Enhancer Activity by Ascl1 and Sox factors

Cristina Minieri

September 2013



Division of Molecular Neurobiology
MRC National Institute for Medical Research
The Ridgeway
Mill Hill, London
NW7 1AA, UK

Cell and Developmental Biology

University College London

A thesis submitted to University College London for the degree of Doctor of
Philosophy

This project has been completed in the laboratory of François Guillemot, in the Division of Molecular Neurobiology at the National Institute for Medical Research, London.

I, Cristina Minieri, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

I would like to thank my supervisor, François Guillemot, for giving me the opportunity to work in his lab and for his guidance and support throughout this project. I would also like to thank Ben Martynoga for advice, discussions and for everything he taught me during this work. I would like to thank all the members of Guillemot's lab for their help, critical discussions, for the pleasant and vibrant environment and for the enjoyable time spent together over these years. I would like to thank in particular: Sebastien Gillotin for his advice and helpful suggestions on this project, but also for being a good friend and for the fun time spent together over these years in our breaks at the canteen, but also at the theatre, at the opera, at the ballet; Debbie van den Berg for critical discussions and technical guidance and suggestions with protein work, and for her "random bursts of curiosity" with the gel shift assays outside the dark room. She really can shift a day of doubts into a day of good hopes for the future; Patricia Garcez, for her lessons of determination and a genuine friendly support, and for the (few) very early mornings in the lab or, even better, at the gym. I would like to thank Surendra Kotecha and Steve Moore for their technical guidance and an introduction to EMSA. I would like to thank my thesis committee Siew-Lan Ang, James Briscoe, and Vassilis Pachnis for critical discussions over the project, with an eye on the route and the other on the captain; the Medical Research Council for funding this project. I am grateful to some lovely supportive people at the NIMR, in particular those who help the students to have a smooth life in the work of every day, Eileen McCarthy, Vicky Millins, and the director of the Studies, Donna Brown. I would particularly like to thank the good friends met at

the NIMR with whom I shared the journey and the experience, Saira Hussain, Emma Cookson, and Jon Pitt.

A special acknowledgement is for those that have always supported me before and beyond this project: all my family; mum and daddy for giving me a multi-angled perspective over things that has guided me to the end between bends and falls; my English family, the Sweeney's, for some lovely and relaxing Bristolian time together.

Finally, I would like to thank Nathan Sweeney, for some corrections on the thesis and for helpful discussions on the project over the years, but mainly for sharing the whole journey, and the much-loved longer one of our lives together.

*«Perché l'animo tuo tanto s'impiglia»,
disse 'l maestro, «che l'andare allenti?
che ti fa ciò che quivi si pispiglia?*

*Vien dietro a me, e lascia dir le genti:
sta come torre ferma, che non crolla
già mai la cima per soffiar di venti;*

Dante, Il Purgatorio – Canto V

Abstract

Transcriptional enhancers and their associated transcription factors (TFs) have a profound impact in gene regulation during development. The TFs Ascl1 and the Sox factors play important roles in many aspects of neurogenesis in vertebrates. Ascl1 induces neuronal differentiation in neural progenitor cells. Sox2 is an essential TF in the maintenance of neural stem cell characteristics, but also functions in the correct specification of neuronal subtypes. SoxC factors have a fundamental role in the establishment of neuronal traits in differentiating neurons. Data previously generated in our lab have identified genomic regions bound by both Ascl1 and Sox2 in close proximity through a ChIP-seq approach, raising questions about possible interactions between these TFs as transcriptional regulators. Here, I have characterised these regions as neural enhancers regulated by Ascl1, Sox2, and SoxC factors using NS5 cells as an *in vitro* model of neural stem cells through luciferase assays. Enhancers activated by overexpression of Ascl1 were classified as neuronal enhancers, since overexpression of this TF induces neuronal differentiation in NS5 cells. Cotransfection experiments and luciferase assays demonstrated that Ascl1 and Sox2 counteract each other in the regulation of the enhancers, whereas Ascl1 and SoxC factors synergistically activate the neuronal enhancers. Mutations of the binding motifs of these TFs and comparisons between the transcriptional activity of the wt and mutant enhancers suggest direct binding of Ascl1 on all the enhancers activated by this TF, but different mechanisms of regulation exist for the Sox factors on the different enhancers analysed. I have shown direct binding of Ascl1 on one of these enhancers, MSB4, by EMSA. In conclusion, I have identified a novel transcriptional regulatory network in the regulation of neuronal differentiation where Ascl1 and SoxC factors synergistically activate neuronal enhancers, and Sox2 counteracts Ascl1 in this activation.

Table of Contents

Acknowledgements.....	3
Abstract.....	5
Table of Contents	6
List of Figures	9
List of Tables.....	10
Abbreviations.....	11
1. Introduction	15
1.1 Development of the Central Nervous System in vertebrates	16
1.1.1 Formation of the Neural Tube	16
1.1.1.1 The anterior-posterior axis	17
1.1.1.2 The dorsal-ventral axis.....	17
1.1.2 Neurogenesis	20
1.1.2.1 Neural progenitor cells of the CNS.....	20
1.1.2.2 Spatial and temporal mechanisms of neural progenitor specification	23
1.1.2.3 Switch from neurogenesis to gliogenesis.....	26
1.1.2.3.1 Radial Glia to astrocytes transition	26
1.1.2.3.2 Generation of oligodendrocytes	26
1.2 Gene regulation by distant-acting enhancers and Gene Regulatory Networks in Embryonic Stem Cells and Neural Stem Cells.....	28
1.2.1 Enhancers as cis-regulatory elements.....	28
1.2.2 TF binding patterns and different models of enhancer activity.....	30
1.2.3 Genome-wide approaches of enhancer identification	32
1.2.3.1 Comparative genomic approaches	32
1.2.3.2 Epigenetic and chromatin marks for enhancer identification	34
1.2.4 Techniques and assays to study enhancer activity and functions.....	36
1.2.4.1 ChIP-chip and ChIP-seq.....	36
1.2.4.2 Transgenic gene reporter assays <i>in vivo</i> and <i>in vitro</i>	37
1.2.4.3 Chromosome Conformation Capture (3C) assay	38
1.2.5 Gene Regulatory Networks in Embryonic Stem Cells and Neural Stem Cells	39
1.3 Proneural proteins and the TF <i>Ascl1</i> in vertebrate neural development.....	42
1.3.1 Biochemical properties.....	42
1.3.2 Cellular functions of proneural proteins and <i>Ascl1</i>	44
1.3.2.1 Notch signalling	44
1.3.2.2 Inhibition of glial fate	46
1.3.2.3 Neuronal subtype specification.....	47
1.3.2.4 Regulation of neuronal migration	49
1.3.3 Molecular mechanisms of proneural genes activity	50
1.3.3.1 Transcriptional targets of the proneural gene <i>Ascl1</i> and its dual role in cell proliferation and neuronal differentiation	50
1.3.3.2 Interaction of the proneural protein <i>Ascl1</i> with other transcription factors.....	52
1.3.4 Proneural proteins and <i>Ascl1</i> in reprogramming.....	53
1.4 Sox factors in vertebrate neural development	54
1.4.1 Biochemical properties of the Sox factors.....	55
1.4.1.1 The HMG domain and its architectural role	55
1.4.1.2 Classification of the Sox proteins	56
1.4.1.3 Sox consensus motif and diverse mechanisms of DNA recognition	57
1.4.1.3.1 Differential affinity of Sox factors for flanking sequences in the DNA recognition.....	58
1.4.1.3.2 Interactions with other TFs in the DNA recognition	58
1.4.1.4 Members of the SoxB2 family act as transcriptional repressor	59
1.4.2 Biological functions of Sox2 in vertebrate neural development	59
1.4.2.1 Role of Sox2 in ESCs and NSCs as stem cell related gene	59

1.4.2.2 Role of Sox2 in neuronal differentiation.....	61
1.4.3 Molecular interactions of Sox2 with other TFs.....	63
1.4.4 Sox2 in reprogramming.....	66
1.4.5 SoxC factors in neural development.....	67
1.4.5.1 Biological functions of the Sox C factors.....	67
1.4.5.2 Molecular characteristics of the SoxC factors.....	68
1.4.5.3 SoxC factors in neuronal differentiation.....	69
1.4.6 Sequential binding of Sox factors in the specification of neural lineage.....	71
1.5 Use of NS5 cells as neural stem cell model system.....	74
1.6 Objectives of the thesis.....	77
 2. Materials and Methods.....	 78
2.1 Cell culture.....	78
2.1.1 NS5 cell culture.....	78
2.1.2 Hek 293T cell culture.....	78
2.2 Chromatin immunoprecipitation (ChIP) and quantitative PCR (ChIP-qPCR).....	78
2.2.1 Chromatin preparation.....	78
2.2.2 Chromatin immunoprecipitation (ChIP).....	79
2.2.3 Quantitative PCR (ChIP-qPCR).....	80
2.3 Plasmids and cloning.....	80
2.3.1 Cloning the enhancer-regions into the luciferase reporter vector.....	80
2.3.2 Site-Directed Mutagenesis.....	81
2.3.3 TFs expression vectors.....	82
2.4 Luciferase assay.....	82
2.5 Electrophoretic Mobility Shift Assay (EMSA).....	83
2.5.1 Protein synthesis.....	83
2.5.1.1 In vitro transcription and translation.....	83
2.5.1.2 Protein expression in 293T cell line.....	84
2.5.1.2.1 Transfection in 293T cells.....	84
2.5.1.2.2 Cell lysis.....	84
2.5.2 Radiolabelling of double stranded oligonucleotides.....	85
2.5.2.1 Annealing of the single stranded oligonucleotides.....	85
2.5.2.2 Radiolabelling reaction.....	85
2.5.3 Gel Shift Assay (EMSA).....	85
2.6 Western Blot.....	86
2.7 Computational analysis.....	87
2.8 Statistical analysis.....	88
2.9 Materials.....	89
 3. Genome-wide identification of neural enhancers regulated by Ascl1 and Sox2..	 96
3.1 Introduction.....	96
3.2 ChIP-seq identification of genomic regions bound by Ascl1 and Sox2: computational analysis (Ben Martynoga's data).....	97
3.3 Selection of putative regulatory elements from Ascl1 and Sox2 co-bound regions identified by ChIP-seq.....	101
3.3.1 Expression pattern in the telencephalon of the genes associated to the bound regions.....	101
3.3.2 Validation of the selected Ascl1 and Sox2 co-bound regions by ChIP-qPCR.....	103
3.3.3 Sequence conservation of the genomic elements.....	106
3.3.4 Recruitment of p300 on the bound regions: a hallmark for the identification of genomic enhancers.....	108
3.4 Transcriptional activity of the genomic regions bound by Ascl1 and Sox2.....	110
3.4.1 Basal activity of the bound regions: regulation by endogenous TFs in NS5 cells.....	111
3.4.2 Regulation of enhancer activity by Ascl1 and Sox2.....	112

3.4.3 Classification of the enhancers: neural stem cells and neuronal enhancers	116
3.5 Regulation of the neural enhancers by Ascl1-EnR and Sox2-EnR	120
3.6 Discussion	121
 4. Identification of the E-boxes and Sox binding motifs and different mechanisms of regulation of the neural enhancers	 126
4.1 Introduction.....	126
4.2 E-boxes and Sox binding motifs in the neural enhancers	126
4.3 Mutations in the sequence of the binding motifs affect enhancer activity	130
4.3.1 Basal transcriptional activity of the mutant enhancers.....	130
4.3.1.1 MSB18 Mutants basal activity.....	130
4.3.1.2 MSB22-short Mutants basal activity.....	131
4.3.1.3 MSB 4 Mutants basal activity	132
4.3.1.4 MSB24-short and MSB11 Mutants basal activity	132
4.3.2 Mutagenesis of the E-box and Sox motifs affects the regulation by exogenous Ascl1 and Sox factors in the neural enhancers	138
4.3.2.1 The neuronal enhancer MSB4.....	138
4.3.2.2 The NSC-active and neuronal enhancer MSB18	139
4.3.2.3 The NSC-active and neuronal enhancer MSB 22-short.....	140
4.3.2.4 The NSC-active enhancer MSB24-short.....	142
4.3.2.5 The NSC-active and neuronal enhancer MSB 11	143
4.4.1 SoxC factors, Sox4 and Sox11, synergise with Ascl1 in the regulation of the neuronal enhancers MSB4, MSB18 and MSB22	148
4.4.2 Dual function of Sox8, member of the Sox E family, in the regulation of the neuronal enhancers MSB 4, MSB 18 and MSB 22	150
4.5 Mutations of E-box and Sox motifs affect the synergy of Ascl1 with SoxC factors, or Sox8 in the activation of neuronal enhancers.....	152
4.5.1 Regulation of MSB4 mutant enhancers by Ascl1 and Sox factors	152
4.5.2 Regulation of MSB22-short mutant enhancers by Ascl1 and Sox factors	155
4.6 Discussion	157
4.6.1 Identification of E-box and Sox motifs in the neural enhancers	157
4.6.2 Mutagenesis of the BMs suggests regulation of the neural enhancers by endogenous bHLH factors and Sox factors	157
4.6.3 Mutagenesis of the BMs suggests different mechanisms of regulation of the neural enhancers by exogenous Ascl1 and Sox2.....	159
4.6.4 Recruitment of different TFs to different BMs in enhancer MSB18	160
4.6.5 SoxC and Sox8 factors in the regulation of the neuronal enhancers and synergy and counteraction with Ascl1	162
4.6.6 Two different mechanisms of synergistic activation of the neuronal enhancers by Ascl1 and the SoxC factors	163
 5. <i>In vitro</i> binding of Ascl1 to the neuronal enhancer MSB4	 167
5.1 Introduction.....	167
5.1.2 Optimisation of the experimental conditions for EMSA experiments	168
5.2 Ascl1 binds <i>in vitro</i> to the E-box in enhancer MSB4.....	170
5.3 Discussion	175
 6. Final discussion and future work.....	 176
References	184

List of Figures

1. Introduction	15
Figure 1.1 Dorsal-ventral specification of the neural tube.....	19
Figure 1.2 Glial nature of Neural Stem Cells in cortical development.	22
Figure 1.3 Spatial mechanisms of progenitor cells specification in the telencephalon....	24
Figure 1.4 Transcriptional regulation by promoters and enhancers.....	29
Figure 1.5 Two models of enhancer activity: the Enhanceosome and the Billboard.....	32
Figure 1.6 Structure of bHLH proteins and dimerization with E proteins.....	43
Figure 1.7 Structure of the Sox factors and the HMG domain.	56
Figure 1.8 The Sox-session.	73
Figure 1.9 Coexpression of Ascl1 and Sox2 TFs in vitro in NS5 cells by immunofluorescence staining.....	76
 3. Genome-wide identification of neural enhancers regulated by Ascl1 and Sox2..	96
Figure 3.1. Computational analysis of Ascl1 and Sox2 ChIP-seq data.....	99
Figure 3.2. Expression pattern of the genes associated to the candidate bound regions.	102
Figure 3.3. Validation of selected regions bound by Ascl1 and Sox2 by ChIP-qPCR...	104
Figure 3.4. Comparison of early and final ChIP-seq data and ChIP-qPCR validation of regions bound by Ascl1 and Sox2.....	105
Figure 3.5. PhastCons diagram of sequence conservation.	108
Figure 3.6. Recruitment of the transcriptional coactivator p300 on the putative regulatory regions.....	110
Figure 3.7. Transcriptional activity of the bound regions in neural stem cells by luciferase assay.....	112
Figure 3.8. Interaction of Ascl1 and Sox2 in the regulation of the neural enhancers. ..	116
Figure 3.9. Classification of neural stem cells and neuronal enhancers.	118
Figure 3.10. Classification of the enhancers activated by exogenous Sox2 or inhibited by exogenous Sox2.....	119
Figure 3.11. Repression of the enhancer activity by Ascl1-EnR and Sox2-EnR.....	121
 4. Identification of the E-boxes and Sox binding motifs and different mechanisms of regulation of the neural enhancers	126
Figure 4.1. E-boxes and Sox motifs within the neural enhancers.....	128
Figure 4.2. Mutagenesis of the E-box and Sox motifs affects basal luciferase activity of the enhancers.....	137
Figure 4.3. Mutagenesis of the E-box and Sox motifs affect the regulation by exogenous Ascl1 and Sox factors in the neural enhancers.....	148
Figure 4.4. Interplay between SoxC or Sox8 factors and Ascl1 in the regulation of the neuronal enhancers.....	151
Figure 4.5. Mutations of E-box and Sox motifs affect the synergy of Ascl1 with SoxC factors, or Sox8 in the activation of neuronal enhancer MSB4.	154
Figure 4.6. Regulation of MSB22-short mutant enhancers by Ascl1 and Sox factors. .	156
 5. <i>In vitro</i> binding of Ascl1 to the neuronal enhancer MSB4	167
Figure 5.1 – Western blot to verify Ascl1 or Sox2 protein expression using either <i>in vitro</i> TnT system or transfection of 293T cells.....	170
Figure 5.2. Ascl1 binds in vitro to the MSB 4 enhancer sequence.	174
 6. Final discussion and future work.....	176

Figure 6.1 Models of interplay between Ascl1 and Sox factors in the regulation of the neuronal enhancers.....	183
--	------------

List of Tables

2. Materials and Methods	78
Table 2.1 ChIP-qPCR primers	89
Table 2.2 Cloning primers.....	90
Table 2.3 Primers for site-directed mutagenesis.....	92
Table 2.4 DNA Oligonucleotides probes for EMSA experiments	93
Table 2.5 Expression vectors used in transfection and luciferase assays	94
Table 2.6 Expression vectors for protein synthesis for EMSA assays	95
3. Genome-wide identification of neural enhancers regulated by Ascl1 and Sox2..	96
Table 3.1. Genomic regions co-bound by Ascl1 and Sox2.....	100
Table 3.2. Sequence conservation and size of the putative regulatory elements cloned in luciferase vector.	107
Table 3.3. Classification of enhancers.....	120
4. Identification of the E-boxes and Sox binding motifs and different mechanisms of regulation of the neural enhancers.....	126
Table 4.1. Binding Motifs consensus sequences within neural enhancers.....	127

Abbreviations

AEA	Anterior Entopeduncular Area
Ascl1	Achaete-Scute Homolog, 1
β-Gal	Beta-Galactosidase
bHLH	Basic Helix-loop-Helix
BM	Binding Motif
BMP	Bone Morphogenetic Protein
BP	Basal Progenitor
Bp	base pair
3C	Chromosome Conformation Capture Assay
CGE	Caudal Ganglionic Eminence
ChIP	Chromatin-Immunoprecipitation
ChIP-chip	Chromatin-Immunoprecipitation followed by hybridization to genomic arrays
ChIP-seq	Chromatin-Immunoprecipitation followed by direct sequencing
CMV	Cytomegalovirus immediate early promoter
CNS	Central Nervous System
CP	Cortical Plate
DCX	Doublecortin (gene)

DNA	Deoxyribonucleic Acid
E	Embryonic Day
EGF	Epidermal Growth factor
ESC	Embryonic Stem Cell
FGF	Fibroblast Growth Factor
GFP	Green Fluorescent Protein
GoF	Gain of Function
GRN	Gene Regulatory Network
IPC	Intermediate Progenitor Cell
iPS	Induced Pluripotent Stem Cell
ISH	In Situ Hybridization
LGE	Lateral Ganglionic Eminence
LoF	Loss of Function
MACs	Model-based Analysis of ChIP-Seq data
Mash1	Mouse Achaete-Scute Homolog, 1
MGE	Medial Ganglionic Eminence
MZ	Mantle Zone
NE	Neural Epithelium
Ngn	Neurogenin

NSC	Neural Stem Cell
NTDs	Neural Tube Defects
ORF	Open Reading Frame
OSVZ	Outer Subventricular Zone
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
RG	Radial Glia
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
SD	Standard Deviation
SDS	Sodiumdodecylsulfate
SEZ	Subependimal Zone
Shh	Sonic Hedgehog
Sox	SRY-related HMG-containing box
SVZ	Subventricular Zone
TAD	Transactivation Domain
TBS-T	Tris-Buffered Saline-Tween
TE	Tris-EDTA
TF	Transcription Factor

TFBS	Transcription Factor Binding Site
TNE	Tris-NaCl-EDTA
TSS	Transcriptional Start Site
VZ	Ventricular Zone
WT	Wild-Type

1. Introduction

Developmental progression and cell fates are established by highly regulated programs of gene expression, which are driven by interconnected gene regulatory networks. The identification of these transcriptional networks regulating stem cell division and differentiation is an essential step to understand how the balance between self-renewal, multipotency and differentiation is achieved. Transcriptional enhancers and their associated Transcription Factors (TFs) have a prominent role in the initiation of gene expression and its regulation.

This thesis focuses on the identification and characterisation of neural enhancers regulated by the TFs *Ascl1* and the Sox factors, in particular Sox2 and the SoxC members, Sox4 and Sox11. These proteins play important roles in many aspects of the neural development and neurogenesis in vertebrates. Neurogenesis is the generation of neurons from neural progenitor cells and is a tightly regulated genetic process controlled by the interplay of many different TFs.

In this introduction, I first give an overview of the development of the Central Nervous System (CNS) with a focus on neurogenesis in vertebrates. Secondly, an overview of enhancers, their functions and their genomic identification is presented. Finally, the role of the proneural protein *Ascl1* and the Sox factors in neural development is discussed.

1.1 Development of the Central Nervous System in vertebrates

1.1.1 Formation of the Neural Tube

Neurulation is the embryonic process that leads to the formation of the neural tube, which will develop into the brain and spinal cord (Gilbert, 2006). It occurs at the end of gastrulation during embryonic development, around E7 in the mouse.

The formation of the neural tube is a complex process in which cells need to change shape, migrate, and differentiate to transform a flat sheet of epithelial cells, the neural plate, into a hollow tube, the neural tube. This process involves morphogenetic events, controlled by distinct molecular pathways.

The closure of the neural tube along the anterior/posterior axis occurs simultaneously in birds where it is initiated at the level of the future midbrain in a “zip-like” manner proceeding in both directions. In mammals, however, it is initiated at several places, usually three sites, along the anterior-posterior axis.

Neural tube closure requires a complex interplay between genetic and environmental factors. Genes such as *Pax3*, *Sonic hedgehog (Shh)* and *openbrain* are essential for the formation of the mammalian neural tube. In addition, dietary factors such as cholesterol and folic acid seem to be of critical importance to the process. The failure of various parts of the neural tube to close leads to neural tube defects (NTDs). Among the most common forms of NTDs are *anencephaly* and *spina bifida*. The first results from a failure of the neural tube to close anteriorly in the cranial region. It is characterised by a partial or total absence of the cranial vault and cerebral hemispheres and is invariably lethal. The latter occurs when the most caudal region of the neural tube fails to close. It

leads to mild or severe lifelong physical and developmental disabilities, depending on how much of the spinal cord remains exposed (Gilbert, 2006).

1.1.1.2 The anterior-posterior axis

The early mammalian tube is a straight structure. However, as the formation of the neural tube proceeds during development, drastic changes occur along the tube. Initially, the neural tube bulges and constricts in three points generating three primary vesicles: the forebrain, midbrain, and hindbrain. Subsequently, the neural tube further subdivides in five secondary vesicles: the telencephalon and the diencephalon, originating from the forebrain; the mesencephalon from the midbrain; and the metencephalon and myelencephalon from the hindbrain. From these primitive vesicles all the structures of the brain and spinal cord will eventually form.

Several signalling molecules control the anterior-posterior patterning of the central nervous system. The combinatorial code of their spatially and temporally distinct concentrations directs the expression of distinct TFs in different defined domains along the anterior-posterior axis. The Hox genes family is among the genes controlling the patterning, defining cell identity along the axis (Gilbert, 2006).

1.1.1.3 The dorsal-ventral axis

The neural tube is also polarised along its dorsal-ventral axis. The dorsal-ventral patterning is established by the expression of bone morphogenetic proteins (BMPs) dorsally, and Sonic hedgehog (Shh) ventrally (Figure 1.1A). BMP4 and BMP7, the dorsalising signals, are expressed and secreted from the roof plate whilst Shh is expressed and secreted from the floor plate. Overlapping gradients of these diffusible molecules induce the coordinated expression of mutually cross-repressive TFs along the

dorsal-ventral axis. For instance, cells adjacent to the floor plate that are exposed to high concentrations of Shh and hardly any BMPs express the TFs Nkx6.1 and Nkx2.2 and become the ventral neurons of the V3 domain (Figure 1.1B). The cells dorsal to this domain, exposed to slightly less Shh and slightly more BMPs instead express the combination of Nkx6.1 and Pax6 TFs. They are the cells of the motor domain, which will become the motor neurons of the ventral spinal cord. The next two groups of cells, exposed progressively to decreasing concentrations of Shh, become the V2 and V1 interneurons domains (Gilbert, 2006) (Figure 1.1B).

Dorsal-ventral patterning in the forebrain is similarly established by Shh and BMPs. Moreover, fibroblast growth factors (FGFs), secreted from the anterior neural ridge and the rostral midline, also contribute to confer ventral identity to the progenitor cells in the forebrain. In the dorsal telencephalon, Wnt-signalling antagonises the ventral identity (Campbell, 2003). The dorsal-ventral patterning of the forebrain establishes two major progenitor domains: the dorsal domain, or pallium, which later forms the cerebral cortex, and the ventral domain, or subpallium, the precursor of the basal ganglia, the amygdala and the septum, and the origin of cortical interneurons. The patterning of the progenitor domains is an important mechanism in development to pre-determine and commit the fate of progenitor cells in the different specific regions of the central nervous system (Gilbert, 2006).

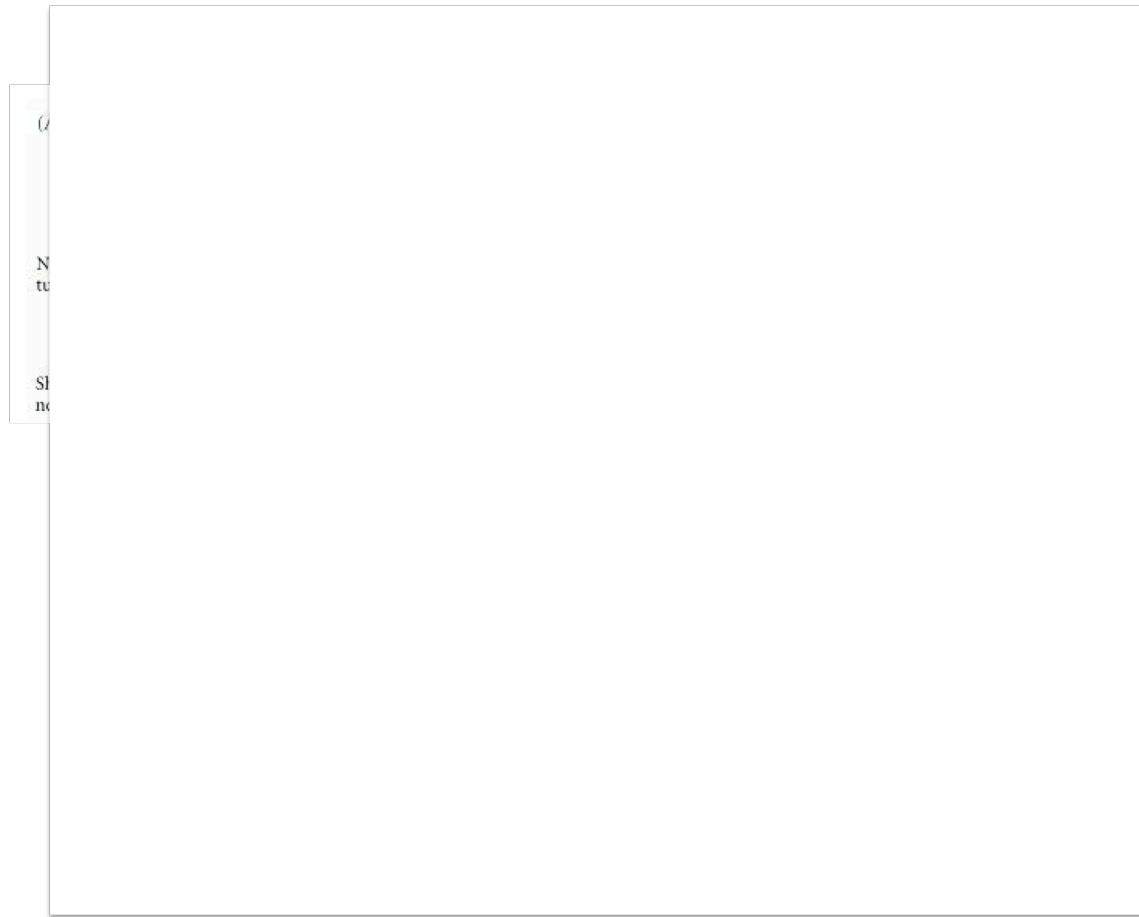


Figure 1.1 Dorsal-ventral specification of the neural tube.

(A) The newly formed neural tube is patterned by two signaling molecules centers. The roof of the neural tube is exposed to BMP4 and BMP7 from the ectoderm; the floor of the neural tube is exposed to Shh from the notochord. Overlapping gradients of these diffusible molecules induce the coordinated expression of mutually cross-repressive TFs along the dorsal-ventral axis. The identity of the neurons in the spinal cord is determined by the combinatorial code of TFs that these cells express in their nuclei and that is established by their position along the dorsal-ventral axis and the exposure to different concentrations of the signalling molecules. Taken from *Gilbert 2006*. **(B)** Specification of neuronal identity in the ventral neural tube by Shh from the floor plate (FP). Relationship between Shh concentrations, the generation of specific progenitor domains and neuronal types *in vitro*, and the distance from the notochord. Taken from (*Wang et al., 2003*).

1.1.2 Neurogenesis

1.1.2.1 Neural progenitor cells of the CNS

The developing mouse cerebral cortex is one of the main model systems for neurogenesis. However, the mechanisms identified in this tissue also operate in a similar way in other regions of the developing brain and spinal cord.

Neurons and glial cells derive from a pseudostratified neuroepithelium of ectodermal origins that line the cerebral ventricles of the developing embryo. The Neuroepithelial (NE) cells undergo rapid symmetric divisions to expand the neural plate.

When cortical neurogenesis begins, around E9-E10 in mouse, the NE cells acquire features of Radial Glial (RG) cells. These cells express glial marker such as GLAST (Glutamate-Aspartate transporter) and BLBP (Brain lipid-binding protein). A series of changes mark the transition from NE cells to radial glial. However, like NE cells, RG cells maintain apical-basal polarity, line the lateral ventricles and undergo interkinetic nuclear migration within the ventricular zone (VZ), therefore maintaining the aspect of pseudostratified epithelium. RG cells persist and represent the principal neural progenitor cells during development of the embryonic CNS. These Neural Stem Cells (NSCs) will give rise to all types of neuronal lineages and macroglial cells, astrocytes and oligodendrocytes, of the brain and the whole CNS (Kriegstein and Alvarez-Buylla, 2009, Martynoga et al., 2012). The progression of lineages from these NSCs occurs so that neurons are generated first, and astrocytes and oligodendrocytes at a later stage (Figure 1.2). During the cortical neurogenesis RG cells undergo asymmetric self-

renewing cell divisions to generate another radial glia cell and a daughter cell that will differentiate into a neuron.

The production of neurons from RG cells is not always a direct step. It can involve the generation of more fate-restricted progenitor cells, called Intermediate Progenitor Cells (IPCs), or Basal Progenitors (BP) (Figure 1.2). It has been showed that basal progenitors (BP) derive from asymmetric divisions of the RG cells (Haubensak et al., 2004, Miyata et al., 2004, Noctor et al., 2004). Therefore, RG are NSCs that either produce neurons directly or through the generation of fate-restricted IPCs, which populate the sub-ventricular zone (SVZ). Thus, the SVZ is a major site of neurogenesis in the developing brain producing neurons away from the ventricle.

IPCs have a multipolar shape, and do not have processes contacting the pial surface or the ventricle. They do not undergo interkinetic nuclear migration. Unlike RG cells, which mainly divide asymmetrically, IPCs undergo symmetric divisions to generate two neurons or two other IPCs (Kriegstein and Alvarez-Buylla, 2009). The number of time IPCs divide to produce more IPCs before generating neurons can vary in different regions of the brain, which further varies between different species. In particular, it has been shown that primate corticogenesis is characterised by the appearance of a large SVZ. The distinguished area of the outer SVZ, the OSVZ, is seat of several mitosis of the IPCs, called outer RGs. These divisions expand the number of mitotic cells that will eventually produce neurons spanning the whole cerebral cortex. Therefore, these IPCs act functionally as transit amplifying cells present in the stem cell lineages of other tissues. Indeed, the expansion of the pool of IPCs producing neurons contributes to the cortical expansion observed in the gyrencephalic (large and highly folded) primate cortex. Conversely, lissencephalic species (with a smooth and non-folded cortex), such as the mouse and most other rodents, also have IPCs in the SVZ, but their number, and

therefore their contribution to the expansion of the cortical plate, is small (Lui et al., 2011). Further supporting the importance of neurogenesis that takes place in the SVZ for the control of cortical size, a congenital disease known as *microcephaly* has been associated to the homozygous silencing of the gene *Tbr2*, a TF considered to be a selective marker for IPCs and functionally required for the SVZ neurogenesis (Sessa et al., 2008, Lui et al., 2011).

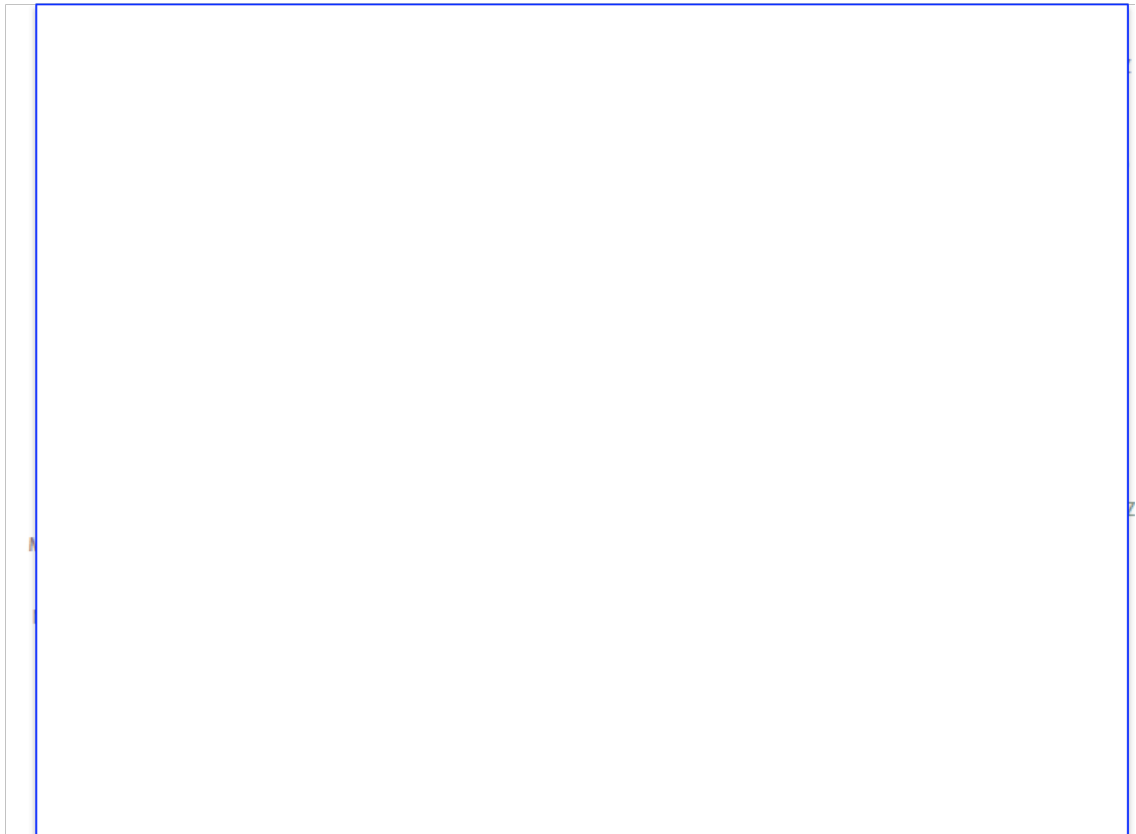


Figure 1.2 Glial nature of Neural Stem Cells in cortical development.

Three modes of neurogenesis: i) directly through asymmetric division; ii) indirectly through generation of neurogenic Intermediate Progenitor Cells (nIPC) and one round of amplification; iii) or again through nIPC but with more rounds of mitosis and amplifications. Radial Glia cells generate also astrocytes and oligodendrocytes at the end of neurogenesis. oIPC: oligodendrocytic IPC; NE: neuroepithelium; MZ: marginal zone; VZ: ventricular zone; SVZ: sub-ventricular zone; IZ: intermediate zone; CP: cortical plate. Taken from *Kriegstein and Alvarez-Buylla, 2009*.

1.1.2.2 Spatial and temporal mechanisms of neural progenitor specification

In the developing forebrain, several signalling centres secrete diffusible molecules that form overlapping gradients and act as morphogens. Sonic hedgehog (Shh) is secreted from the ventral midline of the forebrain; members of the FGF family, such as Fgf8, 15, and 17, are secreted anteriorly by the midline of the telencephalon; members of the Wnt and BMP families are secreted from the medial and caudal area of the cortex. These signalling molecules establish the regionalised expression of homeobox and helix-loop-helix TFs in progenitor cells. The combinatorial code of TFs eventually subdivides the telencephalon in a dorsal domain, or pallium, and a ventral domain, or subpallium (Campbell, 2003). These patterning TFs will induce other TFs, which eventually specify progenitors identity and neuronal fate and phenotypes. Following this principle of regionalisation, the pallium will eventually be subdivided into distinct architectonic and functional areas that later organise the cerebral cortex longitudinally during development.

Collectively, RG cells generate the vast and diversified neuronal subtypes and glial cells of the brain and spinal cord. This is achieved through different signalling and gradients of morphogens, which interact to establish discrete territories of TFs expression, each associated with the production of different types of neurons (Figure 1.3). Thus, segregation of progenitor zones within the forebrain has been characterised, each associated with distinct transcription factor expression. Therefore, RG cells appear to be very heterogeneous in terms of their progenitor functions, depending on the transcription factors they express (Kriegstein and Alvarez-Buylla, 2009). Several studies have focused on the identification of the many TFs involved in the control of neurogenesis, implicated in the proliferation of neural progenitors, and in the control of the tight balance between proliferation and differentiation in the developing brain

(Martynoga et al., 2012). Defects in progenitor divisions and brain growth have been described for mice mutant for the paired homeobox factor Pax6, the homeobox proteins Lhx2 and Arx, the winged-helix protein Foxg1, and the nuclear receptor Tlx (Jacobs and Dinman, 2004, Cahill et al., 1994, Lin et al., 2007, Natesan et al., 1997, Orphanides et al., 1996).

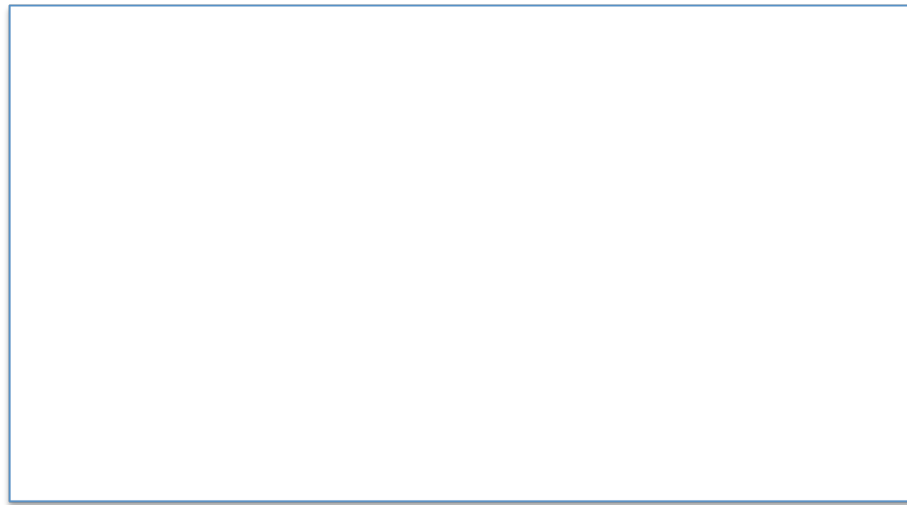


Figure 1.3 Spatial mechanisms of progenitor cells specification in the telencephalon.

Combinatorial codes of TFs expressed in restricted dorsal or ventral domains establish the neuronal fate and identity of progenitor cells in the developing telencephalon. Signaling molecules, such as BMPs, FGFs, and Shh proteins, influence the expression of these different TFs in sharp domains/boundaries. See text for more details. CGE: Caudal Ganglionic Eminence; LGE: Lateral Ganglionic Eminence; MGE: Medial Ganglionic Eminence. Adapted from *Martynoga et al., 2012*.

The basic helix-loop-helix (bHLH) proneural transcription factors are critical regulators of neurogenesis. Among these, the Neurogenins, Ngn1 and Ngn2, and Ascl1/Mash1 are expressed in RG cells of the developing cerebral cortex, already at the time of NE to RG transition (Martynoga et al., 2012). It has been proved that cortical neurogenesis is considerably damaged in embryos deficient for these proneural factors, both *in vivo* and *in vitro* (Nieto et al., 2001). These same proneural factors are also sufficient to induce a full program of neurogenesis. Indeed, overexpression *in vivo* or *in vitro* of either Ngn2

or *Ascl1* can induce a full neuronal differentiation (Mizuguchi et al., 2001, Nakada et al., 2004, Berninger et al., 2007). Moreover, their requirement for neuronal differentiation can be further supported by their reprogramming capabilities. For instance, *Ascl1* can respecify fibroblasts into neurons, with greater efficiency if it is overexpressed with the other TFs *Brn2* and *Myt1l* (Vierbuchen et al., 2010).

Other TFs, such as *Ngn2*, *Insm1* and *AP2 γ* have been implicated in the generation of BPs from RG cells. They all induce *Tbr2*, the specific marker of intermediate progenitors required for their generation (Sessa et al., 2008). Finally, TFs such as *Foxg1* and *Ascl1* control the division of the BPs in the cortex and the ventral telencephalon, respectively, activating cell cycle regulators such as *E2f1* and cyclin-dependent kinases (Siegenthaler et al., 2008, Castro et al., 2011).

In addition to location, temporal patterning in development must determine the different type of neurons generated from neural progenitors. Referring again to the mammalian cerebral cortex as model, this is radially organised in six different layers of neurons, characterised by the expression of distinct molecular markers and projection patterns to and from different areas of the brain. There is a correlation between the location of neurons in different layers and the time of their birth. In particular, neurons in the deepest layers are generated first, and neurons in the upper layers are generated progressively later in development (McConnell, 1995b, McConnell, 1995a).

Dynamic programs of transcription factor expression are involved in this process of specification. For instance, the TFs *Fezf2* and its activated target *Ctip2* have a critical role in the specification of the neurons of the layer V of the cerebral cortex and in the specification of their correct pattern of axonal projections (Arlotta et al., 2005, Molyneaux et al., 2005). *Ctip2* is a specific marker of layer V neurons, so

unsurprisingly the correct specification of both earlier-born and later-born neurons requires the repression of Ctip2. Tbr1 represses Ctip2 for the specification of layer VI neurons, and Satb2 represses Ctip2 for the specification of layer II/III. Therefore, it appears evident that networks of cross-interacting and/or cross-repressing TFs also play a critical role in the temporal specification of neuronal fates from a common progenitor cell in the control of neurogenesis in the developing cerebral cortex.

1.1.2.3 Switch from neurogenesis to gliogenesis

1.1.2.3.1 Radial Glia to astrocytes transition

RG cells persist in the VZ throughout the period of cortical development, as NSCs in neurogenesis and also representing the scaffolding glia for the migration on new-born neurons. At the end of the developmental period, and perinatally in mouse, most radial glia cells lose their attachment to the ventricle and migrate to the cortical plate: this represents a new transformation: the radial glia to astrocytes transition. In this transition, RG cells change from bipolar or unipolar cells with only a radial process to a multipolar shape without radial processes acquiring astrocytic morphology. Some astrocytes might undergo symmetric divisions locally to amplify their number before terminal differentiation, therefore representing a population of IPCs with astrocytic fate. In some vertebrate species, the transition of RG to astrocytes does not occur, and RG cells persist postnatally as glia cells (Kriegstein and Alvarez-Buylla, 2009).

1.1.2.3.2 Generation of oligodendrocytes

RG generate oligodendrocytes in addition to neurons during development. Oligodendrocytes originate in different locations in three distinct developmental waves. The earliest wave originates in the VZ from Nkx2.1-expressing progenitor cells of the medial ganglionic eminence (MGE) and the anterior entopeduncular area (AEA) around

E12 in mouse and entering the developing cortex around E16. A second wave follows, originating from Gsh2-expressing progenitor cells of the lateral and/or caudal ganglionic eminence (LGE/CGE) around E16-E18. Finally a third wave originates postnatally from Emx1-expressing progenitors from the dorsal cortex itself (Kessaris et al., 2006).

The origin of the oligodendrocytes from RG confirm the idea of these cells as the NSCs generating all neurons and glia cells of the entire CNS (Kriegstein and Alvarez-Buylla, 2009).

1.2 Gene regulation by distant-acting enhancers and Gene Regulatory Networks in Embryonic Stem Cells and Neural Stem Cells

1.2.1 Enhancers as cis-regulatory elements

Gene expression is regulated through the integrated activity of many regulatory elements, including promoters and promoter proximal elements, and elements that are located at a greater distance from Transcriptional Start Sites (TSS), such as enhancers, silencers and insulators. Enhancers and their transcription factors have a prominent role in the initiation and transcriptional regulation of gene expression.

Transcriptional enhancers are non-coding regulatory sequences that activate promoter-mediated transcription of their target genes. They can be located at great distances, tens to hundreds of thousands of base pairs away from their target genes. Moreover, they can be located upstream, downstream or even within their target gene and can modulate its expression independently of their orientation in respect to the TSS direction. Enhancer sequences are characterised by a dense clustered aggregation of Transcription Factor Binding Sites (TFBSs). Upon the right occupancy of TFs and their interplay, recruitment of transcriptional coactivators and chromatin remodelling proteins occur on the enhancers. There, the protein complexes are thought to facilitate DNA looping, which brings the protein complexes on the enhancer into close physical proximity to those on the promoter (Figure 1.4). This conformational change allows promoter-mediated gene activation (Visel et al., 2009b).



Figure 1.4 Transcriptional regulation by promoters and enhancers.

General transcription factors (green ovals) bind to core promoter regions through recognition of elements such as the TATA boxes and Initiators (INR). Sequence-specific TFs (red trapezoid) bind cis-elements (dark blue box) in the proximal promoter region and stabilize the recruitment of the transcriptional machinery through direct interaction with general transcription factors on the promoter. Promoter activity is further stimulated to higher levels of transcriptional activity (represented by ++) by enhancer elements (light blue box), which can be located at far distance from the promoter region. Sequence-specific TFs (orange octagon) bind enhancer elements and can stimulate transcription by recruiting a histone-modifying enzyme (for example, a histone acetyltransferase (HAT)) to create a more favourable chromatin environment for transcription (for example, by histone acetylation (Ac)). Upon binding of TFs, enhancers relocate in physical proximity with the promoter through DNA looping and activate promoter-mediated transcription. Adapted from *Farnham, 2009*.

Some characteristics of the enhancers, such as their capability to activate the expression of their target genes independently of their location and orientation in respect to the TSS, were already described when the first enhancer was discovered over thirty years ago (Banerji et al., 1981). Banerji and colleagues reported that the expression of a β -globin gene could be strongly enhanced by a 72 bp repeated sequence element of the viral SV40 DNA when this element was included in the β -globin -plasmid recombinants transfected into HeLa cells. The viral DNA element could enhance the expression of the β -globin gene independently from its orientation and at many positions from the gene

TSS including being located either 1400 bp upstream or 3300 bp downstream of the TSS. The element was named an enhancer as it is capable of enhancing the expression of a gene. A reorganisation of the chromatin structure by the enhancer element was already hypothesised to be the possible mechanism of regulation of the gene expression (Banerji et al., 1981).

1.2.2 TF binding patterns and different models of enhancer activity

Individual enhancers are small sections of DNA, typically a few hundred bp long, that serve as operational platforms to recruit TFs. TFs recognise and bind specific DNA sequences on the enhancers, the TFBSs or motifs, to regulate transcription and gene expression. Different combinations of several TFs can bind the enhancers in different spatial domains of expression in a developing embryo or at different developmental stages. This combinatorial binding can result in discrete and precise patterns of transcriptional activity (Spitz and Furlong, 2012).

The organisation of the TFBSs within an enhancer in terms of motif composition and motif positioning is referred to as enhancer architecture, and these properties have been investigated in detail leading to the establishment of different models of enhancer activity. In some cases motif positioning, that is the relative order, orientation and spacing of TF motifs within an enhancer, appear as strict sequence constraints of the TFBSs or different modules within the regulatory elements. A very well studied example of this enhancer architecture model is the interferon- β enhancer, which led to the “enhanceosome” model of enhancer activity (Panne et al., 2007, Panne, 2008) (Figure 1.5A). In this model 8 TFs bind to a DNA sequence of 55bp in a highly cooperative manner and small changes within the regulatory sequence are sufficient to alter binding of all TFs and eliminate enhancer activity. The TFs recruited form a highly

ordered protein interface that requires a strict positioning of TF binding motifs relative to each other along the DNA. The overlapping of TF binding sites creates a composite element that ensures that the enhancer operates as a single unit of regulation, in an on/off switch manner. The sharp switch-like activation of this enhanceosome is strictly dependent from the tight organisation and positioning of TFBSs within the regulatory element. Enhancers like the interferon- β do not display functional redundancy, and indeed the interferon- β enhancer sequence is well conserved.

However, most developmental enhancers do not conform to this model, displaying a much more relaxed and independent recruitment of TFs to autonomous modules within the enhancer sequence. This allows a more flexible positioning of motifs and has led to the “billboard” model of enhancer activity, in which enhancers behave as “information display” elements where distinct TFs might bind either cooperatively or independently from one another with only few constraints in the enhancer architecture of the binding motifs (Arnosti and Kulkarni, 2005) (Figure 1.5B). It is worth noting that the enhanceosome and the billboard models of enhancer activity represent the two extreme ends of a continuum spectrum of enhancer architectural diversity.

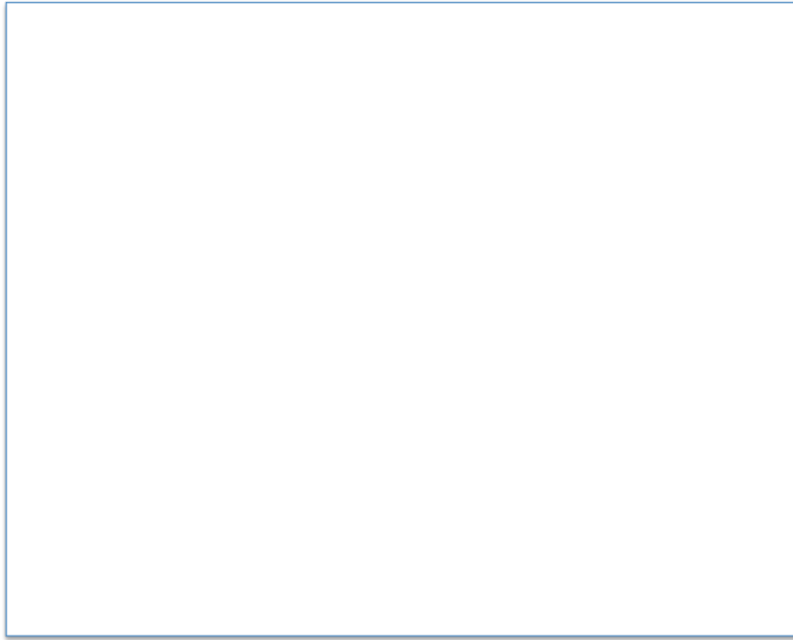


Figure 1.5 Two models of enhancer activity: the Enhanceosome and the Billboard.

(A) In the enhanceosome model, TFBSs within the enhancer sequence are strictly constrained in term of motif positioning. This is to allow a high cooperative assembly and interaction between all the TFs (coloured ovals in the cartoon) recruited on the regulatory element, leading to gene activation. Disruption or displacement of even one only TFBS or loss of one TF binding causes the enhancer to be inactive and transcription to be turned off. (B) In the Billboard model, the enhancer contains independent activating or repressing modules that can regulate gene expression independently from one another. Independent TFs can activate or repress the activity of the enhancer when they interact with the transcriptional machinery (coloured ovals represent activator, and red square represent repressor). Adapted from *Arnosti and Kulkarni, 2005*.

1.2.3 Genome-wide approaches of enhancer identification

1.2.3.1 Comparative genomic approaches

An extensive identification of enhancers was initially achieved through comparative genomic studies. These approaches are based on the notion that genomic regulatory sequences are under negative evolutionary selection as changes in functional sequences might have negative consequences (Visel et al., 2009b, Dickel et al., 2013). Therefore, conservation in non-coding sequences could be the first way to identify potential

enhancers throughout the genome of different species. Thus, analysis and statistical measures of evolutionary sequence constraint provide a tool to identify potential enhancers within the big portion of non-coding sequence in the genomes of vertebrates (Visel et al., 2009b).

Pennacchio et al. give an example of the application of this approach in a work, where a large set of ancient human-to-fish conserved and human-to-rodents ultra-conserved elements were computationally filtered and identified as putative enhancers and a subset of them were tested for their activity *in vivo* in transgenic mouse enhancer assays performed at E11.5 (Pennacchio et al., 2006). Briefly, in this assay the genomic conserved fragment is linked to a minimal promoter fused to a *lac-Z* reporter gene. The transgenic construct is transiently transfected into the developing embryo. Following to this, β -Gal staining of the whole-embryo is performed to assess the spatial enhancer activity of the cloned element *in vivo*. 29% of ancient human-to-fish conserved, and 61% of human-to-rodents ultra-conserved elements tested were validated as active enhancers *in vivo*. The entire enhancer dataset from this study was submitted to the VISTA enhancer browser. This is a database of tissue-specific human enhancers identified through comparative genomics and experimentally validated through *in vivo lac-Z* transgenic mouse enhancer assays (Visel et al., 2007).

Although comparative genomics led to successful identification of genome-wide enhancers, it has some limitations. For instance, conservation cannot predict when and where enhancers are active in development or an adult organism, always requiring experimental validations of their identification that is hard to achieve on a genome-wide scale (Visel et al., 2009b).

1.2.3.2 Epigenetic and chromatin marks for enhancer identification

As a strategy alternative and complementary to comparative genomics, the generation of genome-wide maps of specific chromatin marks powerfully allows the identification of enhancers and other regulatory elements. This approach became possible thanks to the improved development of techniques such as ChIP-ChIP and ChIP-seq, which allow the generation of a genome-wide binding profile of DNA binding proteins and chromatin remodelling complexes enriched at specific regulatory elements. A description of these techniques is given in the next paragraph 1.2.4.

Several studies provided insights of the epigenetic marks found at specific categories of regulatory elements, allowing a precise discrimination among them. Relevant to the identification of enhancers in particular, monomethylation at lysine residue 4 of histone H3 (H3K4me1) was found to mark enhancers and specifically discriminate them from promoters marked by tri-methylation at the same position (H3K4me3). Moreover, acetylation at lysine 27 of histone H3 (H3K27ac) has also been associated to enhancers and more precisely it distinguishes active enhancers from poised elements containing H3K4me1 alone. (Heintzman et al., 2007, Heintzman et al., 2009, Creighton et al., 2010). Interestingly, in a study by Mikkelsen et al., ChIP-seq of different epigenetic marks recruited on promoters allowed the generation of “chromatin-state maps” for pluripotent and lineage committed mouse cells where active, poised or repressed promoters could be discriminated at the different developmental cell states according to the chromatin-state and marks recruited (Mikkelsen et al., 2007). The study shows that developmental programs from pluripotent to terminally differentiated cells driven by change in gene expression and transcriptional competence are paralleled by characteristic chromatin-state changes. This work by Mikkelsen et al was also one of the first examples of the application of the ChIP-seq to locate and identify genome-wide

enhancers and to generate chromatin maps of different cell types in a multicellular organism.

Also the acetyltransferase and transcriptional coactivator p300 is a protein associated to the enhancers. In a work by Visel et al., the genome-wide distribution of this protein was determined by ChIP-seq directly from developing mouse tissues, in particular the forebrain, midbrain and limb. This study is particularly worth noticing as the recruitment of the protein p300 was determined *in vivo* from developing tissues, rather than in cell culture, but more importantly because the tissue-specific occupancy of p300 could in most cases accurately predict the *in vivo* pattern of expression driven by these enhancers, providing an important advantage over comparative genomic methods for enhancer identification (Visel et al., 2009a). To further prove the strength of this method over comparative genomics to identify active enhancers, the same research group reported the identification of weakly conserved heart enhancers that escaped comparative approach for their identification, employing the same p300 ChIP-seq approach from developmental heart tissue (Blow et al., 2010).

All together the studies presented here represent the first maps of histone marks and protein p300 and dataset available to identify and predict the activity of transcriptional enhancers throughout the genome. Despite the importance and the progress made in the identification and characterisation of the non-coding genome in the regulation of gene expression, *in vivo* and *in vitro* biological studies are still necessary to understand the function and the mechanisms of regulation by non-coding enhancers and the interplay of TFs involved in their activation. An overview of the current techniques available to study enhancer activity is given in the next paragraph.

1.2.4 Techniques and assays to study enhancer activity and functions

1.2.4.1 ChIP-chip and ChIP-seq

The methods of ChIP-chip and ChIP-seq have allowed the identification of enhancers on a genomic scale through detecting the binding of chromatin marks and cofactors on the regulatory sequences. These methods are based on traditional ChIP experiments with different downstream analysis. Briefly, ChIP involves crosslinking of DNA-binding proteins with the DNA by treating cells with formaldehyde and shearing chromatin usually by sonication. An immunoprecipitation of the cross-linked chromatin is performed using an antibody against the specific TF or protein of interest. This results in the identification of all the binding sites in the genome for the protein of interest. After reversal of the crosslink and purification of the precipitated chromatin fragments the sample can be analysed by PCR to study particular gene. However, genome-wide analysis can be performed by microarray (ChIP-chip) or sequencing (ChIP-seq). For ChIP-chip the immunoprecipitated sample and the input chromatin, as a control, are labeled with fluorescent dyes and are hybridized to microarrays. For the ChIP-seq instead the immunoprecipitated DNA fragments are sequenced through next-generation sequencing techniques, and computationally mapped to a reference genome. The results from ChIP-seq are based on statistical analysis of read counts and advanced computational ChIP-seq analysis tools are available to identify ChIP-seq peaks. The ChIP-seq analysis covers the entire mappable portion of the reference genome without need to restrict the analysis to its subregions, as it happens instead in the ChIP-chip by hybridization to microarrays. Mainly this improvement has contributed to the success of

the ChIP-seq as method of choice compared with the ChIP-chip, together with lower cost of the experiments and an unambiguous identification of the peaks (Farnham P. J., 2009, Visel et al., 2009b).

Although techniques like the ChIP-chip and ChIP-seq are very successfully for the identification of enhancer elements genome-wide, they make use of indirect properties of the regulatory elements (such as the occurrence of chromatin marks or binding of specific TFs of interest) rather than assessing their functionality and activity.

1.2.4.2 Transgenic gene reporter assays *in vivo* and *in vitro*

Transgenic reporter assays provide a more functional approach to test enhancers, by addressing their ability to drive gene expression. In a typical transgenic reporter assay a putative enhancer fragment is cloned upstream of a reporter gene driven by a promoter, which has a minimal or no activity by itself, but that respond to the input of the adjacent enhancer. In this way the activity of the enhancer is revealed by the expression of the reporter gene.

In vivo transgenic mouse reporter assays are one of the most used techniques to detect enhancer activity *in vivo*. In these experiments the enhancer to test is linked to the reporter gene, typically *LacZ*, and then the transgene construct is delivered into mouse zygotes through pronuclear injection. The resulting transgenic embryo will be tested for β -gal activity to visualise the expression pattern and *in vivo* activity of the enhancer element in the embryonic tissues.

In vivo transgenic mouse assay can't be used for quantitative analysis of the enhancer activity nor to detect modest alteration to enhancer intensity or quantitative effects of enhancer mutations. These effects and a quantitative measure of enhancer activity have

been studied predominantly using *in vitro* reporter assays, where the enhancer is coupled to a *luciferase* reporter gene and transiently transfected into cells. The transgenic reporter gene intensity can be measured quantitatively with the use of a luminometer (Rosenthal, 1987, Naylor, 1999, Schenborn and Groskreutz, 1999).

The luciferase assay can allow quantification of the gene expression driven by an enhancer element in different cell lines for instance to test the specificity of the enhancer in different cell context that might be similar or very different according to the choice of the experimental systems and the purposes of the comparison. In the same way, the effect of the input of specific TFs activity on the enhancer can be measured by gain or loss of function of the TF of interest or by manipulating the amount of the same TF. Finally, effects of mutations in the enhancer sequence, for instance in the consensus motifs of known TFBSs, can be detected and quantified. For these applications, the luciferase assay persists as one of the best techniques to study and assess the biological function of enhancers and to characterise the functions of the several modules that makes the enhancers and the effect of specific TFs recognising and binding to the TFBSs. Despite the advances of large-scale genome-wide techniques to identify and characterise enhancers *in vivo*, luciferase assay experiments are still needed in parallel to these approaches to understand the biological function of enhancers, and the mechanisms of their activation and regulation by TFs. The work presented in this thesis demonstrates the advantage of the application of this technique.

1.2.4.3 Chromosome Conformation Capture (3C) assay

A big remaining challenge in the study of enhancers is to determine the relationship between enhancers and their target genes. Although comparisons between ChIP-chip or ChIP-seq with transcriptome data from microarray and RNA-seq can give clues in the

association enhancer-target gene, they can't give evidence of a direct interaction enhancer-promoter and therefore enhancer-target gene which would be necessary in the establishment of gene regulatory networks on a genomic scale. Current views support the idea that most enhancers establish direct physical interactions with their target gene promoters (Symmons and Spitz, 2013). These interactions can be detected by chromosome conformation capture assay (3C) and its derivative technologies. This technique is based on a biochemical strategy that allows representation of the 3D organisation of the DNA and the chromosome topology. Subsequent steps of fixation, digestion, and re-ligation of fixed chromatin followed by quantification of the ligation junctions allows obtaining a one dimension cast of the 3D chromatin structure (Zhu et al., 2013, de Wit and de Laat, 2012). Overall, the main picture emerging from studies based on these approaches is that both promoters and enhancers are frequently engaged in multiple interactions so that enhancer-promoter interactions are not exclusive. For most genes, the elements that regulate their expression will be found in *cis*, although at distances that could be hundreds of kilobases. However, enhancers have been reported to act also promiscuously, activating neighbouring but biologically irrelevant genes (Symmons and Spitz, 2013). In conclusion, these studies are pointing out that many current approaches taken to associate enhancers with their target genes, such as the previously mentioned comparison of ChIP-seq data with transcriptome data, might be misleading and new approaches need to be developed.

1.2.5 Gene Regulatory Networks in Embryonic Stem Cells and Neural Stem Cells

I have discussed in the previous paragraph that enhancers are characterised by a dense cluster of TFBSs and that the right occupancy of TFs and their interplay on the

enhancers is an essential step in the control of gene regulation. In accordance with this notion, several studies in the past few years have generated binding profiles of TFs genome-wide and have highlighted that many of them co-bind and co-regulate a common set of targets in the establishment of transcriptional gene regulatory network (GRN) that underpin the cell state and its maintenance.

To gain insight into the GRN in embryonic stem cells (ESCs) for instance, Chen and co-workers have used ChIP-seq to map the binding of 13 TFs with relevant roles in ES biology as key regulators of pluripotency and self-renewal (Chen et al., 2008). Remarkably, this study has identified two clusters of TFs. One cluster includes the TFs Nanog, Sox2, Oct4, Smad1, and STAT3 that tend to co-occur and co-bind quite often a common set of target genes. Likewise, the second cluster includes n-Myc, c-Myc, E2f1, and Zfx, which co-bind a second set of distinct targets. The first cluster with Oct4 binds to MTLs (Multiple TFs binding Loci), genomic regions with high dense TFBSs spots, which exhibit characteristics of enhancers. Indeed, they recruit the transcriptional coactivator p300 together with the TFs listed; they are associated with H3K4me3 mark, and show specific luciferase activity in ES cells. The high dense cluster of TFBSs within relative compact genomic segments, and the requirement of the TF Oct4 as key regulator for the recruitment of Smad1 and Stat3 suggest that the MTLs identified have characteristics of enhanceosomes. Interestingly, ES cell specific gene expression is associated with binding of many of the factors studied. Precisely, the cluster with Oct4, Sox2, Nanog, Smad1 and STAT3 is thought to be mainly involved in the regulation of the pluripotency genes, while the Zfx, E2f1, Myc cluster often regulates genes involved in self-renewal, including the promotion of cell division. Therefore, based on an association between TFs binding and gene expression, the authors of this study have

constructed a GRN model in ESCs, which underpins the features of pluripotency and self-renewal typical of this cell state.

The work of Chen and colleagues has identified Oct4, Sox2 and Nanog as key players in the cluster of TFs underpinning pluripotency in the mouse ESCs. Interestingly, this is consistent with the findings of a previous work by Boyer and co-workers, which pose the same three TFs at the top of the GRN controlling pluripotency and self-renewal in human ES cells (Boyer et al., 2005). In a genome-scale location analysis using ChIP-chip, Boyer et al. revealed that these three TFs co-occupy the promoters of a large population of genes, usually other downstream TFs, involved in the maintenance of the stem cell state. Oct4, Sox2, and Nanog regulate also transcription of their own genes and therefore establish autoregulatory and feedforward loops circuitry contributing to pluripotency and self-renewal in the human ESCs (Boyer et al., 2005). The mechanisms of cooperativity and interactions between Oct4 and Sox2 in the regulation of their target genes are discussed in more details in paragraph 1.4.3 “Molecular interaction of Sox2 with other TFS”.

In more recent years there have been studies attempting to establish transcriptional gene regulatory networks also in neural stem cells (NSCs). For instance, a work by Southall and Brand in *Drosophila* has identified a common set of target genes involved in the self-renewal and differentiation of the NSCs co-regulated by the TFs Asense, one of the *Ascl1* ortholog in *Drosophila*, Deadpan, Snail, and Prospero (Southall and Brand, 2009). This study unveiled NSC transcriptional networks essential for the balance between self-renewal and differentiation.

The TFs Ascl1 and the Sox factors, in particular Sox2, are involved in the regulation of several transcriptional networks and developmental programs in the NSCs in cooperation with other TF-partners.

In the next paragraphs, I give an overview of the biochemical characteristics and biological functions of Ascl1 and the Sox factors in neural development.

1.3 Proneural proteins and the TF Ascl1 in vertebrate neural development

1.3.1 Biochemical properties

Proneural proteins are basic helix-loop-helix (bHLH) transcription factors. Many bHLH TFs have important role in the generation and differentiation of tissues throughout the all animal kingdom. Proneural proteins, in particular, have a prominent role to initiate the development of neuronal lineages and to promote the differentiation of committed neural progenitor cells. The TF Ascl1, also known as Mash1 in mouse, is among these proneural proteins, and its role in neural development is described in the next paragraphs.

This big class of TFs is characterised by the presence of the bHLH domain, the structural motif responsible for their DNA-binding and dimerization properties (Murre et al., 1989) (Figure 1.6). Proneural proteins, like all other bHLH TFs, bind the DNA as heterodimers that are formed with ubiquitously expressed bHLH proteins, the E proteins, such as E12 and E47 in mammals. The bHLH heterodimers are formed by

interactions of the two helices of each (TF) partner to form a four-helix bundle (Figure 1.6).

The class of the bHLH TFs, including the proneural proteins, specifically bind a core hexa-nucleotide DNA sequence motif, CANNTG, called E-box. The basic region contacts the main groove of the DNA and confers sequence-specificity of the binding, while the HLH region is responsible for the dimerization with other bHLH proteins (Massari and Murre, 2000).

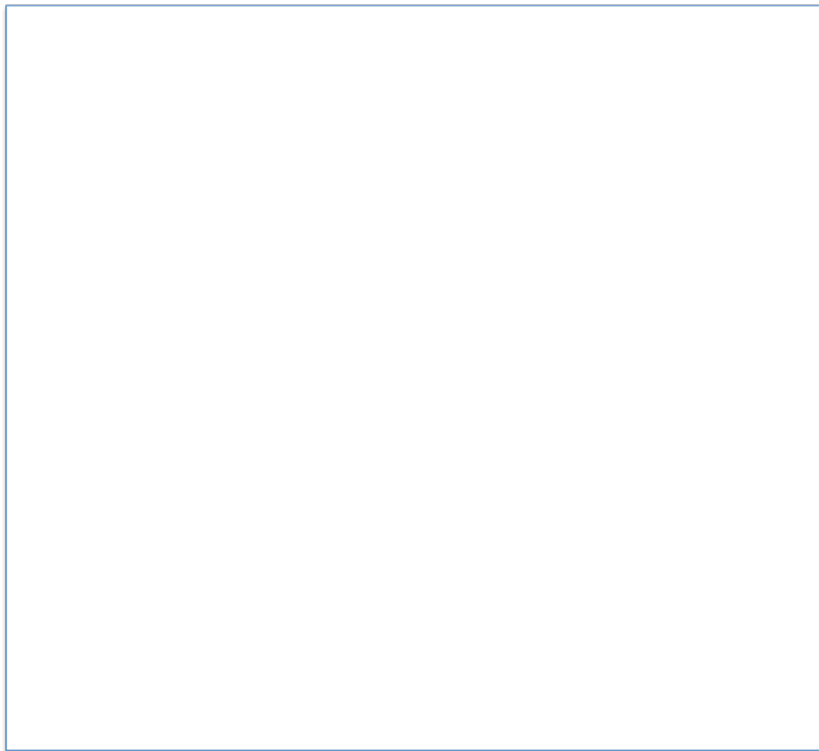


Figure 1.6 Structure of bHLH proteins and dimerization with E proteins.

(A) Schematic representation of the structure of a bHLH dimer bound to the DNA. The basic region binds to the major groove of the DNA while the two α -helices of both partners in the dimer form a four-helix bundle. (B) bHLH proteins bind the DNA as heterodimers with the E proteins, which are ubiquitously expressed bHLH proteins. Adapted from *Bertrand et al., 2002*.

Proneural proteins and most bHLH TFs act mainly as transcriptional activators. In the case of the Neurogenins, Ngns, transcriptional activation is induced through recruitment of the transcriptional coactivator and acetyltransferase p300 (Koyano-Nakagawa et al.,

1999, Sun et al., 2001, Lee et al., 2009). Identification of the cofactors that interact with other proneural proteins, such as *Ascl1*, is indispensable to understand their mechanisms of transcriptional regulation.

1.3.2 Cellular functions of proneural proteins and *Ascl1*

Proneural proteins are expressed almost exclusively in progenitor cells where they specify neuronal fate and initiate programmes of differentiation. Since this thesis is focused on the transcription factor *Ascl1* as transcriptional regulator in neural stem cells, and since the developing cerebral cortex was presented as main model for neurogenesis in the previous paragraphs, this section describes the functions of proneural proteins, in particular *Ascl1*, mainly in the telencephalon where *Ascl1* and the two Neurogenins, *Ngn1* and *Ngn2*, are the only proneural proteins expressed.

Loss of function (LoF) and Gain of function (GoF) studies of proneural proteins *in vivo* and in neural stem cultures *in vitro* revealed the biological functions of this class of bHLH TFs in neural development.

1.3.2.1 Notch signalling

The activation of the Notch signalling pathway is one of the earliest functions of proneural proteins in neural development. This proneural function is well conserved in the development of the nervous system from *Drosophila* to vertebrates (Casarosa et al., 1999, Bertrand et al., 2002).

Both proneural proteins *Ascl1* and *Ngn2* have been shown to induce the expression of *Dll1*, the Notch ligand, in progenitors expressing the proneural factors (Castro et al.,

2006). The expression of *Dll1* induces Notch signalling in the neighbouring cells preventing them from differentiating, in a process known as “lateral inhibition”.

Upon activation of the signalling, the transmembrane protein Notch is cleaved and releases the Notch intracellular domain (NICD), which moves to the nucleus, where it forms a complex with the DNA-binding protein RBPj. The NICD-RBPj complex then induces the expression of the bHLH repressor factors Hes1 and Hes5, the Notch effectors, which repress the expression of proneural genes. Thus, proneural genes inhibit their own expression in adjacent cells. The Notch signaling is therefore very important to maintain the pool of progenitor cells. In the absence of this pathway, all neural progenitors differentiate prematurely altering the balance between differentiating and proliferative stem cells and leading to depletion of the full spectrum of diverse cells born at later stages. An analysis of *Ascl1* mutant mice in the ventral telencephalon revealed a loss of expression of the *Dll1*, *Dll3*, and *Hes5* genes. This study proved the role of *Ascl1* as important regulator of neurogenesis in the ventral telencephalon, where it specifies neural precursor and controls the timing of their specification (Casarosa et al., 1999).

The Notch signaling is also responsible for the typical salt and pepper expression pattern of the proneural genes. In particular, a recent new view of the Notch signaling proposed by Kageyama and colleagues confers the salt and pepper expression pattern of *Ngn2* in cortical VZ cells to oscillating expression of Hes1, causing a consequent oscillating expression of *Ngn2* and its target *Dll1* in an opposite phase, in a dynamic process of lateral inhibition (Kageyama et al., 2008, Shimojo et al., 2008).

1.3.2.2 Inhibition of glial fate

Proneural genes such as *Neurogenins* and *Mash1* are expressed almost exclusively in progenitor cells in the ventricular zone of the telencephalon since E8.5 in mouse (Guillemot and Joyner, 1993, Sommer et al., 1996). They also are expressed almost exclusively in a complementary manner, with Neurogenins expressed at high level exclusively in the dorsal telencephalon, and *Ascl1* expressed at high level in the ventral telencephalon, and at low level in the dorsal telencephalon (Fode et al., 2000). This restricted and mainly non-overlapping patterns of expression of *Ascl1* and *Ngns* is also suggestive of their role in the specification of distinct neuronal identities.

Indeed, *Ascl1* null mutant embryos present defects in neurogenesis particularly in the MGE of the ventral telencephalon. Discrete neuronal populations of the basal ganglia and cerebral cortex, originated by progenitors located in the MGE, are subsequently affected. This study proved the role of *Ascl1* as a determinant gene for the specification of neuronal identity in the ventral telencephalon (Casarosa et al., 1999).

In the same way, *Ngn2* single and *Ngn1*; *Ngn2* double mutants present defects in the neurogenesis of the dorsal telencephalon, where populations of early born lower-layers cortical neurons lose expression of dorsal-specific markers, such as *Tbr1*, *Math2*, *Nscl1* and *Nscl2* (Fode et al., 2000). The analysis of *Ngns* mutant mice proved the role of this proneural factor in the specification of cortical neurons.

Interestingly, loss of both *Ngn2* and *Ascl1* in *Ngn2*; *Ascl1* double mutant mice resulted in a severely affected neurogenesis in the developing cerebral cortex with premature onset of astroglial generation. Importantly, *in vitro* cultures obtained by *Ngn2*; *Ascl1*

double mutant cortical progenitors proved that either Ngn2 or Ascl1 is required in the radial glial cells of the cerebral cortex to maintain their neurogenic potential and suppress a premature activation of the astrocytic programme. Therefore, inhibition of the glial fate in an early stage of cortical neurogenesis is one of the main functions of the proneural genes Ascl1 and Ngns (Nieto et al., 2001).

1.3.2.3 Neuronal subtype specification

I have already mentioned that the proneural genes Ngns and Ascl1 have another important function in the vertebrate neurogenesis, which is the neuronal subtype specification of the neural progenitors, with Neurogenins imparting a dorsal cortical phenotype and Ascl1 specifying ventral identity of precursor and the subsequent generation of specific neurons with different molecular characteristics.

In particular, analysis of neurogenin2 mutant mice has proved that this proneural factor is critical for the specification of glutamatergic cortical projection neurons of the lower layers of the cortex. In contrast, specification of cortical projection neurons in the upper layers appears to be independent from Ngn2 activity (Schuurmans et al., 2004). A microarray analysis of the Ngn1; Ngn2 double mutants showed that many targets of these proneural genes include TFs specifically involved in the determination of a glutamatergic cortical projection neuron lineage such as Neurod1, Neurod2, Neurod6, and Tbr2 and also in the expression of the vesicular glutamate transporters vGlut1 and vGlut2 (Schuurmans et al., 2004). Therefore Ngns are the proneural factors responsible for the establishment of a glutamatergic differentiation programme in cortical progenitors.

In parallel, studies of *Ascl1* mutant have proved its role in the specification of ventrally generated basal ganglia neurons and cortical interneurons (Casarosa et al., 1999). The ventral telencephalon is also the domain of highest expression of *Ascl1*, only expressed at low level in the VZ of the dorsal telencephalon. Interestingly, *Ngns* mutants exhibit an overexpression of *Ascl1* in the cortical neural progenitors and a subsequent ectopic expression of ventral markers such as *Dlx1*, *Dlx2*, *Dlx5*, and *GAD67*, a gene encoding an enzyme for the synthesis of the GABA neurotransmitter (Fode et al., 2000). Therefore, Fode and colleagues showed that *Ngns* function is required also to repress dorsal expression of *Ascl1* and the establishment of a ventral GABAergic differentiation programme, and to maintain the two separate expression domains of the proneural genes. In this study the phenotype of a double mutant *Ngns2; Ascl1* was also analysed, and the use of a mouse line with a replacement of *Ngns* by *Ascl1* in dorsal progenitor was employed to dissect the role of proneural factors in neuronal subtype specification in the telencephalon. Analysis of these mutants demonstrated that ectopic expression of *Ascl1* is required and sufficient to confer a GABAergic ventral identity to neurons in the developing cerebral cortex. Indeed, the expression of ventral markers is lost in *Ascl1; Ngns2* double mutants showing that is the ectopic expression of *Ascl1* to confer this ventral phenotype.

To further confirm the role of *Ascl1* in the specification of GABAergic telencephalic neurons, Berninger and colleagues demonstrated that overexpression of *Ascl1* in neural stem cell cultures or cortical astrocyte cultures generates neurons with molecular and electrophysiological characteristics of GABAergic neurons (Berninger et al., 2007, Heinrich et al., 2010). Altogether, the studies here presented illustrated that the specification of neuronal subtype is another important function of the proneural proteins.

1.3.2.4 Regulation of neuronal migration

The proneural proteins Ngn2 and Ascl1 have been implicated in the regulation of neuronal migration in the embryonic cerebral cortex. Impaired neuronal migration has been described in the cortex of single progenitor cells where the proneural functions were lost by Sh-RNA-mediated silencing or electroporation of Cre-recombinase in conditional mutant embryos (Heng et al., 2008, Pacary et al., 2011).

At molecular level, both proneural proteins Ngn2 and Ascl1 control neuronal migration through down-regulation of RhoA-signalling. This is achieved by activation of *Rnd* genes, a family of a small group of atypical GTP-binding molecules. Ngn2 directly activates Rnd2, and Ascl1 directly activates Rnd3. Sh-RNA mediated knock-down of Rnd proteins also exhibited neuronal migration defects. Both Rnd proteins then achieve regulation of neuronal migration through down-regulation of RhoA-signalling. Rho-A signalling can control neuronal migration due to its role in regulating the actin cytoskeleton.

Indeed, as Heng and colleagues and Pacary and colleagues have shown, aberrant phenotypes of neuronal migration in cortical neurons mutant for expression of the proneural proteins can be rescued by suppression of the RhoA-signalling either by restoring the expression of Rnd2 and Rnd3 (Heng et al., 2008, Pacary et al., 2011) or by down-regulation of RhoA-signalling via RhoA-ShRNA (Pacary et al., 2011).

1.3.3 Molecular mechanisms of proneural genes activity

1.3.3.1 Transcriptional targets of the proneural gene *Ascl1* and its dual role in cell proliferation and neuronal differentiation

The characterisation of the direct transcriptional targets of the proneural genes provides in-depth understanding of their functions and of the programmes of neurogenesis they regulate. A work by Castro and colleagues characterised on a genome-wide scale the transcriptional programme regulated by *Ascl1* in the embryonic brain by combining location analysis of *Ascl1*-bound sites in developing telencephalon and neural stem cell cultures with expression profiling of genes deregulated by overexpression or deletion of the proneural gene (Castro et al., 2011).

The results of this analysis showed that *Ascl1* regulates a large number of target genes with diverse biological functions, including for instance genes that control early steps of lateral inhibition, cell fate commitment and specification, and later steps of neuronal differentiation, such as synthesis of the neurotransmitter and axonal morphogenesis. Also the molecular functions of the target genes are very diverse, as *Ascl1* can broadly activate genes whose functions are transcriptional activity, signal transduction, transporter activity and cytoskeletal activity. In conclusion, *Ascl1* can directly control all phases (early and late) and aspects of the neurogenic program.

Surprisingly, the analysis also pointed out that *Ascl1* can activate a large number of positive cell cycle regulator genes, involved in the progression of the cell cycle and in mitosis, such as *E2f1*, *Cdk1*, *Cdk2*, *Skp2*, *Cdc25b*. Indeed, loss of function analysis in embryo and in NS cell cultures confirmed that *Ascl1* is required for normal progenitor divisions. As expected, target genes involved in the cell cycle arrest, a known function

of *Ascl1*, were also identified, including new and previously identified targets, such as *Fbxw7*, *Gadd45g*, *Ccng2*, *Hipk2*, and *Prmt2*. Thus, a new dual role emerged for *Ascl1* in neurogenesis as the same TF can control opposite aspects of the cell cycle, its progression and its withdrawal, and link together cell proliferation and expansion of neural progenitors and their cell cycle exit and neuronal differentiation. Evidence related to the spatial and temporal pattern of expression of the cell cycle arrest genes suggest that they are activated by *Ascl1* at a later phase of neurogenesis.

It is surprising that the same TF can activate the two opposite processes in the same lineage. Interestingly, also a work by Southall and Brand has reached the same conclusion for *Asense*, one of the *Ascl1* ortholog in *Drosophila*, which can promote neuroblast self-renewal, besides its known role in inhibiting cell divisions of neuroblast daughter cells and its pro-differentiation role (Southall and Brand, 2009). This suggests that the dual role of *Ascl1* might be evolutionary conserved. As the authors of the paper pointed out, the advantage of having the same TF activating genes with opposite roles in the regulation of the cell cycle might be a rapid switch from a progenitor state to a neuronal differentiated state ensuring that the two programmes underpinning these two states are mutually exclusive (Castro et al., 2011).

In conclusion, a new uncharacterised function of *Ascl1* as regulator of genes associated with a neural stem cell undifferentiated state has emerged. The mechanisms by which the same TF *Ascl1* can select and activate the two different set of targets in subsequent phases are unknown. The formation of complexes in which *Ascl1* interacts with different TFs and cofactors can be one possibility.

1.3.3.2 Interaction of the proneural protein Ascl1 with other transcription factors

Transcription factors regulate the expression of target genes as part of large transcriptional protein complexes, which include both other TFs and cofactors. Given the big diversity of target genes and programs of differentiation regulated by the proneural factors, it is possible that interactions with different TFs, occurring at subsequent phases of the developmental process or in different cell compartment, might be responsible for this selection of target genes. It is known that interactions between TFs play an essential role in this selection and account for the specificity towards a different subset of targets at different times.

A work by Castro et al., has shown that the regulation of the gene encoding the Notch ligand, Delta1, involves cooperative binding of Ascl1 and the Brn factors (Brn1 and Brn2 (Alvarez-Bolado et al., 1995)) to an evolutionary conserved motif in this gene (Castro et al., 2006). The study demonstrated that Ascl1 and the Brn factors recognize and bind (both *in vivo* and *in vitro*) adjacent motifs on the enhancer of the *Delta1* gene, respectively an E-box and an octamer, and that this interaction is required for the activation of the enhancer and its function. The cooperative binding reflects recruitment by Pou protein of Ascl1 to the E-box of the Delta1 enhancer. Moreover, Castro et al. have shown that this cooperation can occur also to regulate the expression of other target genes besides Delta1. By searching for conserved Ascl1-Brn-like motif in the whole genome and by performing ChIP with chromatin prepared from embryonic telencephalon, they identified other genes associated with this conserved consensus sequence, and therefore candidate targets of a synergistic regulation by Ascl1 and the Brn proteins. These genes are involved in Notch signaling, cell cycle progression, cell differentiation, and cell migration (Castro et al., 2006). This study provides an example of interaction of the proneural factor Ascl1 with another family of TFs to control the

regulation of gene expression and access to a specific subset of targets, coordinating different programs of neurogenesis.

1.3.4 Proneural proteins and Ascl1 in reprogramming

Recent reprogramming studies are highlighting the essential role of fate determinant transcription factors in the specification of differentiated cell types and in the transcriptional program they control, besides their therapeutic potential and applications. For instance, the well-known role of the bHLH factors Ascl1 and Neurog2 as neurogenic factors in the specification of neuronal lineages has been further strengthened by these studies.

Initially, studies where neural stem cell cultures expanded *in vitro* (of adult SEZ origin) or cortical astrocytic cultures could be induced to generate fully differentiated and functionally synapse-forming neurons from forced expression of both proneural proteins Ascl1 and Neurog2, via retrovirus mediated transduction, had confirmed their neurogenic potential, as previously mentioned elsewhere in this introduction (Berninger et al., 2007, Heinrich et al., 2010). These studies have shown how the expression of these single factors alone was sufficient to confer the neuronal lineages. Later on, combinatorial expression of specific TFs was proved to generate fully differentiated and functional neurons from mouse embryonic fibroblasts (MEFs) and hepatocytes, therefore cells with an endodermal embryonic origin rather than ectodermal. In particular, the combination of Ascl1, Brn2, and Myt1l, named as BAM pool of genes, was identified as essential gene regulatory network capable of inducing a transcriptional program typical of functional neurons, and surprisingly to fully silence and “repress” the donor/origin cells transcriptome. The molecular mechanisms underpinning these reprogramming processes are not known. As the authors of these studies suggest, a new

gene regulatory network must be established downstream of strong cell-fate-determining TFs that strengthens the induced transcriptional program. Changes in the transcriptional activity might determine also reorganisation of chromatin and epigenetic features in term of DNA methylation, histone modification and chromatin remodelling complexes, which stabilise the new transcriptional network. Therefore the importance of interactions between different TFs and cofactors in the establishment of GRNs and determination of cell fate, which was discussed in the previous paragraphs, is further highlighted in these reprogramming studies (Vierbuchen et al., 2010, Marro et al., 2011).

1.4 Sox factors in vertebrate neural development

Transcription factors of the Sox family are well-established players of the stem cell state and cell fate decision during development. Sox factors take part in many developmental processes indeed. Several situations have been reported where many Sox proteins are simultaneously expressed, as in the case of Sox2, Sox9, and Sox21 in the NS5 cells (Martynoga, unpublished data) or the SoxB1 and SoxB2 members *in vivo* (Bylund et al., 2003, Sandberg et al., 2005). In some cases they act redundantly, in some others they are antagonistic or involved in different aspects of the developmental process (Bylund et al., 2003, Sandberg et al., 2005). In this thesis, I mainly focus on the biological and molecular functions of Sox2 and the SoxC members, in particular Sox4 and Sox11, in neural development because these factors are the transcriptional regulators involved with Ascl1 in the regulation of the enhancers identified in this study.

1.4.1 Biochemical properties of the Sox factors

1.4.1.1 The HMG domain and its architectural role

The first Sox gene identified was the mammalian testis-determining factor *Sry* (Gubbay et al., 1990, Sinclair et al., 1990), which opened the study on the family of the Sox factors (Sry box containing factors). Like the Sry protein, all Sox factors carry the characteristic HMG domain, which is responsible for binding the DNA. This domain consists of three α -helices arranged in an “L” shape and an N-terminal beta strand (Werner et al., 1995, Wegner, 1999, Lefebvre et al., 2007) (Figure 1.7). Unlike most binding domains of other TFs, the L-shaped HMG domain binds the DNA in the minor groove and induces a significant bend of the of the DNA helix. Thus, HMG domain-containing proteins can act as architectural proteins as they can alter the conformation of the DNA and increase its protein accessibility. In this unique feature lies the ability of the HMG-containing proteins in recruiting other TFs on the DNA, also factors binding on non-adjacent sites, to form active transcriptional complexes (Wolffe, 1994, Werner and Burley, 1997, Wegner, 1999, Lefebvre et al., 2007).

Most Sox proteins have also other functional domains besides the HMG domain. These domains are generally highly conserved among members of the same group, whereas they are different among proteins from distinct groups. Among these domains there are the transactivation (TA or TAD), transrepression, and dimerization domains (Lefebvre et al., 2007).

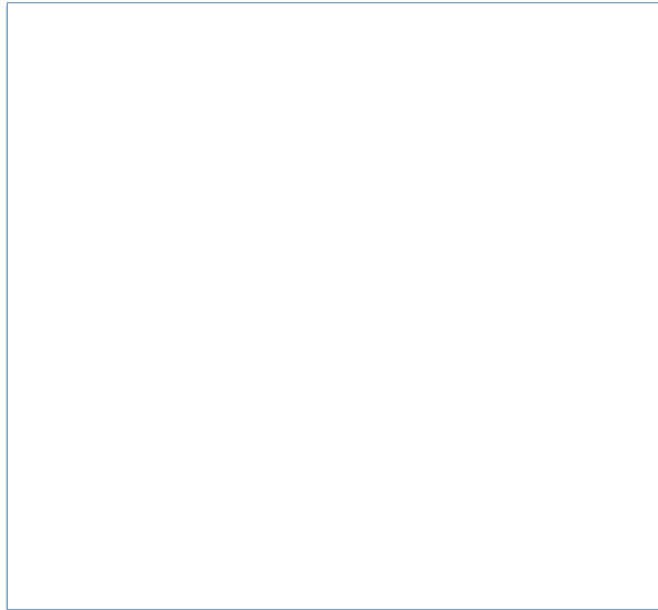


Figure 1.7 Structure of the Sox factors and the HMG domain.

(A) Structure of the mouse Sox proteins and their functional domains. Sox proteins are represented as boxes. Black box represents the HMG domain, which binds the DNA; transactivation domain (TA) represented as vertical stripes, in the C-terminal portion of the protein. Position of the first and last amino acid residues and boundaries of functional domains are indicated. (B) Schematic representation of the HMG domain bound to the DNA. The domain folds into an L shape composed of three α -helices. It fits into the minor groove of the DNA and bends it creating an angle varying between 30° and 110° . Two HMG domains are shown binding on the DNA in opposite orientation. The α -helices are represented as cylinders. N-terminal and C-terminal sequences to the HMG domain and sequences linking the helices are represented as thin rods. The DNA bending angle is also illustrated in the figure above the HMG domains. Adapted from *Lefebvre et al., 2007*.

1.4.1.2 Classification of the Sox proteins

Proteins containing an HMG domain with 50% or higher amino acid (aa) similarity to the sequence of the HMG domain of Sry are classified as Sox proteins. Up to date, 20 Sox genes have been identified (Schepers et al., 2002). Moreover, two Sox-like genes have been found also in the unicellular organism *Monosiga brevicollis*, suggesting the ancient origin of the Sox proteins (Guth and Wegner, 2008).

Sox proteins are divided in groups or subfamilies termed A to H and the classification is based purely on sequence comparisons (Schepers et al., 2002). Proteins of the same group share usually more than 80% aa sequence similarity in their HMG domain (see

Figure 1.7 for representation of the structure of the Sox factors). Moreover, proteins in the same group have similar biochemical properties and thus tend to act redundantly when coexpressed, whereas proteins from different groups usually perform different functions, despite recognising the same DNA consensus motif (reviewed in Wegner 2010) (Wegner, 2010).

1.4.1.3 Sox consensus motif and diverse mechanisms of DNA recognition

All Sox proteins bind to a common consensus motif 5'-(A/T)(A/T)CAA(A/T)G-3' (Harley et al., 1994). This heptameric motif is short and degenerate and fails to discriminate between different Sox proteins. However, Sox proteins differentially bind and recognise their target genes through different binding sites suggesting that the mechanism of recognition does not rely exclusively on the sequence of the consensus motif. Target gene selectivity by different Sox proteins can be achieved and influenced by differential affinity for particular flanking sequences next to the Sox consensus sites, interaction with other cofactors and TFs, homo and heterodimerization with other Sox proteins, post-translational modifications of the Sox factors (reviewed in Wegner 2010). These mechanisms allow recruitment of Sox factors on many and very different binding sites throughout the cell genome, which probably account for the very different molecular and biological functions that the same Sox factor can exhibit in different biological context. Furthermore, these mechanisms also allow regulatory regions to discriminate and recruit a specific Sox factors rather than others. The two mechanisms relevant to the topic of this thesis are discussed in the two following paragraphs.

1.4.1.3.1 Differential affinity of Sox factors for flanking sequences in the DNA recognition

Sox proteins show preferences for specific bases within the sequence flanking the consensus motif with an effect on the selection of the binding sites. For instances, binding affinity for binding sites where some specific nucleotides were flanking the Sox sites were identified for Sox9, Sox10 and Sox4, as briefly reviewed in Wegner 2010. Also different preferences for bases within the heptameric consensus help to discriminate binding sites. Different affinity in the binding of Sox4 to primary and secondary motifs with two different nucleotides in the sequence were analysed in a work by Jauch and colleagues in a novel approach aiming to elucidate the effect of alternative motifs in the binding profile of TFs in the genome (Jauch et al., 2012).

1.4.1.3.2 Interactions with other TFs in the DNA recognition

Sox factors are known to interact with a high number of different TFs and it is believed that these interactions can help in the selection of the binding sites. Very often Sox binding sites are found in the immediate vicinity of binding sites for other TFs in a conformation termed composite elements. Thus, protein-protein interaction might determine which Sox factor is recruited on the composite element and also influence the affinity of a specific Sox partner for that binding site in a cooperative binding between the two TFs on the composite element. The best characterised examples of composite elements recruiting Sox2 factors are the Sox2-Pax6 interaction in the regulation of the lens-specific *δ-crystallin* enhancer (Kamachi et al., 2001) and Sox2-Oct4 composite element identified in the regulatory regions of many ESCs genes, such as *Fgf4*, *Nanog* and some others (Yuan et al., 1995, Rodda et al., 2005, Remenyi et al., 2003). The

biological implications of these molecular interactions are discussed in paragraph 1.4.3 “Molecular interaction of Sox2 with other TFs”.

1.4.1.4 Members of the SoxB2 family act as transcriptional repressor

Although most Sox genes act as transcriptional activators, Sox14 and Sox21, members of the SoxB2 subgroup, are known to act as transcriptional repressors, in contrast to their close related factors of the SoxB1 subgroup (Uchikawa et al., 1999). They have a transrepression domain in their C-terminal portion of the protein. A work by Sandberg et al. has shown that Sox21, acting downstream of proneural proteins, counteracts the functions of the coexpressed SoxB1 members to induce neurogenesis in the chick developing neural tube. Sox21 competes with SoxB1 proteins for binding the same target gene promoters and repressing their activity (Sandberg et al., 2005).

1.4.2 Biological functions of Sox2 in vertebrate neural development

1.4.2.1 Role of Sox2 in ESCs and NSCs as stem cell related gene

Sox2 is active in the embryonic nervous system from the early stages of development. It is expressed in mouse already in the presumptive neuroectoderm (Avilion et al., 2003). It is indeed essential to the formation of the neuroectoderm and acts as a neural competence factor, as shown by studies in *Xenopus* where microinjection of dominant negative version of Sox2 (Sox2dn) suppresses the expression of regional neural markers and inhibits the neural differentiation of the ectoderm (Kishi et al., 2000). In the early neuroectoderm, all three members of the SoxB1 group, Sox1, Sox2, and Sox3 proteins, are expressed in an overlapping manner and act redundantly probably due to their

biochemical and functional similarities. Therefore they compensate for the loss of each other and do not show evident phenotypes in the single mutant of each factor, as reviewed in Wegner and Stolt (Wegner and Stolt, 2005).

All three SoxB1 factors universally mark neural progenitor stem cells in the whole CNS. In chick, constitutive expression of Sox2 inhibits neuronal differentiation and maintains progenitors in a stem cell state. In contrast, inhibition of Sox2, mediated by Sox2EnR electroporation in the developing spinal cord, causes the loss of neural progenitors from the VZ, cell cycle exit and early onset of neuronal differentiation, as demonstrated in a work by Graham and colleagues (Graham et al., 2003).

Particularly relevant to the work of this thesis is a study by Bylund et al. showing that vertebrate neurogenesis, initiated by the activity of the bHLH proneural proteins, is counteracted *in vivo* by the activity of the SoxB1 TFs, Sox1, Sox2, and Sox3, which show overlapping expression pattern and redundant functions (Bylund et al., 2003). The work of Bylund and colleagues has found that Sox1-3 factors keep neural cells undifferentiated, as overexpression of these factors in chick neural tube by *in ovo* electroporation is associated to a lack of expression of pan-neuronal markers such as p27^{kip1}, NeuN, or Tuj1 in the marginal zone of transfected cells. Therefore Sox1-3 prevent cells from upregulating post-mitotic neuronal markers. Interestingly, Sox1-3 factors do not repress the expression of the bHLH proneural factors but rather repress differentiation events downstream of these. Conversely, proneural proteins induce neurogenesis by suppression of Sox1-3 expression in neural progenitors, a critical step in the acquisition of a neuronal fate. I have already mentioned in the previous paragraph 1.4.1.4 that the repressor activity of Sox21 can induce neurogenesis possibly by repressing the activation of the same target genes of Sox1-3, downstream of the proneural proteins (Sandberg et al., 2005). In conclusion, *in vivo* studies from Jonas

Muhr lab in the chick developing neural tube have shown how critical is the balance between Sox1-3 and Sox21 factors in neural progenitors either to keep them in an undifferentiated state or to promote neurogenesis possibly through control of the same target genes. An important activity of the proneural proteins in their induction of neurogenesis seems to be the upregulation of Sox21 as essential step towards the suppression of Sox1-3 activity and commitment of progenitor cells to the neurogenic program. Together, these studies have therefore identified the critical interplay between bHLH proneural proteins and the SoxB1 and SoxB2 factors in the establishment of the neurogenesis *in vivo* (Bylund et al., 2003, Sandberg et al., 2005).

1.4.2.2 Role of Sox2 in neuronal differentiation

Interestingly recent gene targeting studies investigated the function of Sox2 in two neuronal systems: the developing eye and brain. These studies uncovered a role for Sox2 not only in the maintenance of neural stem cells, but also in the differentiation of specific neuron sub-types (Pevny and Nicolis, 2010).

Neural stem cells from the brain (telencephalon) of Sox2 hypomorphic mutant mice show a specific defect in neuronal differentiation, with comparatively normal self-renewal (Cavallaro et al., 2008). Sox2-deficient neural stem cell cultures generate normal numbers of beta-tubulin-positive cells, yet these are poorly arborized, and are negative for mature markers such as MAP2, NeuN, and Calretinin (a marker of GABAergic differentiation). The *in vivo* counterpart to this includes reductions in cell number and arborisation, and the delayed migration of GABAergic interneurons from two areas: cortical interneurons originating from the embryonic ganglionic eminences, and olfactory bulb interneurons, generated postnatally. The differentiation defect could be rescued, *in vitro*, by a Sox2-expressing lentivirus at early (proliferating) but not late

(postmitotic) stages of differentiation. Sox2-deficient differentiating cells coexpress markers of neuronal (beta-tubulin) and glial (GFAP) differentiation, contrary to wild type cultures. Ectopic Sox2 expression by the lentivirus represses endogenous GFAP expression, and the GFAP gene is a direct target of Sox2. Thus, Sox2 acts in neuronal progenitors to downregulate genes of an alternative (glial) differentiation fate (Cavallaro et al., 2008).

Parallel study of complete Sox2 ablation and hypomorphic mutations in the retina also demonstrated profound defects in neuronal differentiation (Taranova et al., 2006). Indeed, while complete Sox2 ablation leads to a major loss of neural progenitors and terminally differentiated cells downstream to them, hypomorphic/null heterozygotes revealed specific differentiation defects in one important type of neuron, the retinal ganglion cell (RGC), whereas other neuronal types remain comparatively normal, although they are mislocalized. RGCs (positive for the mature markers Neurofilament, Brn3b, Islet1) are indeed absent in mutant postnatal retina: young mutant RGC are initially born during embryogenesis (as recognized by positivity for beta-tubulin and Brn3b), but fail to terminally mature, reach their appropriate location, or to appropriately develop axons that enter the optic nerve (Taranova et al., 2006).

In these studies the use of mutant allelic series to decrease/abolish Sox2 expression led to the recognition of a marked dosage-dependence of Sox2 function. Furthermore the effects of Sox2 deficiency are variable and context and species dependent. These are important aspects to consider when assessing the role of Sox2 in the regulation of its target genes. Complex interactions of Sox2 with other regulatory factors, as well as possible redundancy with other SoxB1 genes, can affect the outcome of Sox2 deficiency and its role in gene regulation (Pevny and Nicolis, 2010).

Several interactions of Sox2 with different cofactors have been reported, and, as for Ascl1, both roles in NSC maintenance and neuronal differentiation have been reported. This might give rise to multiprotein-DNA complexes with specific functional properties, where the recruitment of different cofactors results in different transcriptional outcome. Besides TFs and cofactors recruitment, chromatin-modifying enzymes may also be important partners in Sox2 function (Pevny and Nicolis, 2010).

Interactions of Sox2 with other TFs are discussed in the next paragraph, where some examples of the best-characterised and more relevant interplay of factors have been selected and presented.

1.4.3 Molecular interactions of Sox2 with other TFs

Sox2 interacts with many TFs, whose binding sites are often in the vicinity of the Sox site. There are evidence supporting a model in which the transcriptional regulatory function of Sox2 is dependent on cooperation with another TF and the establishment of a pair Sox2-partner (Yuan et al., 1995, Kamachi et al., 2001, Kondoh and Kamachi, 2010). The selection of the partner in the pair depends from the DNA sequence flanking the Sox site, and the availability of the partner factor in a specific cellular context (Kondoh and Kamachi, 2010). The biological consequences of this mechanism of regulatory function by Sox2, and all Sox factors more in general, is that each Sox-partner TF pair appears to select a specific set of regulatory target genes. The resulting gene expression patterns characterise a particular cell state and subsequent differentiation. For instance, Sox2-Oct4 pair (Oct4 was previously named Oct3; from here the name Oct3/4 sometimes used in this thesis and in the referenced literature) selects and regulates the expression of a set of genes active in the ESCs, necessary to the establishment of the stem cell state, such as the previously mentioned *Fgf4*, *Nanog*,

Utl1, *Fbx15* and the same *Sox2* and *Pou5f1*, the gene encoding the Oct4 protein (Yuan et al., 1995, Rodda et al., 2005, Chew et al., 2005, Remenyi et al., 2003, Tokuzawa et al., 2003). In the NSCs instead, the alternative pair Sox2-Brn2 activates the *Nestin* enhancer, and also a *Sox2* enhancer (SRR2 enhancer) regulating the expression of this gene specifically in the NSCs of the telencephalon (Tanaka et al., 2004, Miyagi et al., 2004, Miyagi et al., 2006). Alternatively, the pair Sox2-Pax6 selects and regulates other lens-related genes together with the previously mentioned DC5 enhancer of the *δ -crystallin* gene, therefore inducing initiation of the lens differentiation process (Kondoh and Kamachi, 2010, Kamachi et al., 2001). Moreover, a wide range of different Sox factors and different partners could be mentioned and involved in the specification of different cell types, or in the activation of targets in different cellular context. For instance, the pair Sox11-Brn2 regulates the activity of the previously mentioned *Nestin* enhancer specifically in the SVZ of the developing spinal cord, thus replacing the pair Sox2-Brn2 that regulates the activity of the same enhancer exclusively in the VZ of the developing spinal cord (Tanaka et al., 2004). Thus, when any component of the Sox-partner TF pair is exchanged with alternative factors, either the Sox factor or its partner, a transition of the gene expression occurs in a cell, leading to progression of the developmental processes.

The Sox2-Oct3/4 pair also activates their own genes in a co-regulatory loop, so that the cell state is maintained and stabilised. A major enhancer of *Pou5f1* gene (the gene encoding the Oct4 protein) is activated by the Sox2-Oct3/4 pair in the ES cells, and the same pair also activates a *Sox2* enhancer regulating the expression of *Sox2* in the same stem cell context (Chew et al., 2005, Okumura-Nakanishi et al., 2005, Masui et al., 2007, Tomioka et al., 2002).

Two main molecular mechanisms underlie Sox-partner TF interactions: cooperative DNA binding; and protein interactions (dependent upon DNA binding), which bring to the formation of potent transactivation complexes. Sox2 and Oct4 bind cooperatively to the enhancer of the *Pou5f1*, *Utf1* and *Fbx15* genes to regulate their expression, for instance, so that the affinity of one factor for the binding site is enhanced by the binding of the other factor (Remenyi et al., 2003, Tokuzawa et al., 2003). Moreover, protein-protein interactions induce the formation of potent transactivation complexes required for the activation of the target genes. In the case of the activation of the *Nestin* enhancer in neural stem cells for instance, Sox2 and Brn2 can establish protein-protein interaction that are independent from DNA binding leading to the synergistic activation of the regulated enhancer (Tanaka et al., 2004). The activation of the DC5 enhancer of the δ -*crystallin* gene in the lens development requires instead cooperative binding of Sox2 and Pax6 to the DNA sequence, which is indispensable for the formation of a potent transactivation complex and the establishment of protein-protein interaction between the two TFs (Kamachi et al., 2001). The work of Kamachi and colleagues has demonstrated that Pax6 undergoes Sox2-interaction dependent and DNA binding sequence-dependent conformational changes that are necessary to the formation of the active transactivation complex and to the synergistic activation of the δ -*crystallin* enhancer by the two TFs. The configuration of the binding motifs in the sequence appears to be indispensable for the conformational changes of Pax6, as demonstrated by insertion of bases in the sequence sufficient to the loss of the transactivation potential (Kamachi et al., 2001).

The prominent role of Oct4 in the maintenance of the pluripotent state has been well established. However, Oct4 can also specify and regulate the primitive endodermal lineage (Niwa et al., 2000). Aksoy and colleagues have demonstrated that Oct4 can regulate endodermal specification by an alternative partnering with Sox17 rather than

Sox2 on an alternative “compressed” Oct-Sox motif associated with endodermal genes (Aksoy et al., 2013, AlFatah Mansour and Hanna, 2013). This is in contrast with the recruitment of the Oct4-Sox2 pair on the “canonical” Oct-Sox motif identified in the regulation of the pluripotency genes in ESCs. Therefore, Oct4 can drive alternative developmental programs of pluripotent or endodermal fate by shuffling Sox partners and selecting alternative enhancers. This study is a further example of the alternative pairs of TFs and members of the Sox family for the selection of binding sites and regulatory circuits that drive differentiation during development.

Another important aspect in the gene regulation by Sox2 is the possible interaction with chromatin remodelling complexes. Interestingly, Engelen and co-workers have demonstrated that Sox2 interacts directly with Chd7, a chromatin remodelling ATPase complex (Engelen et al., 2011). Chd7 was identified as interactor of Sox2 through mass spectrometry, and their direct physical interaction was confirmed by co-immunoprecipitation. The two factors share many binding sites genome-wide, as proved by ChIP-seq experiments, and regulate a set of common target genes, as proved by knockdown of each factor and microarray analysis.

1.4.4 Sox2 in reprogramming

The role of Sox2 and its interplay with Oct4 and Nanog in the establishment and maintenance of the ESCs characteristics has been previously discussed in this thesis. Takahashi and Yamanaka have shown that the combination of Sox2 with Oct4, c-Myc, and Klf4 is essential to reprogram mouse embryonic and adult fibroblast into induced Pluripotent Stem cells (iPS), ES-like cells with morphology, properties, and expression of cell marker genes typical of the ESCs (Takahashi and Yamanaka, 2006).

Relevant to this thesis, a work by Karow and colleagues shows that the combination of Ascl1 and Sox2 can induce reprogramming of human pericytes derived from the adult cerebral cortex into fully functional neurons (Karow et al., 2012). Pericytes are a cell type involved in the establishment and maintenance of the blood-brain barrier and regulation of the local blood flow (Armulik et al., 2011). Remarkably, pericyte cultures used in the work by Karow and colleagues lack completely expression of neuronal markers such as β III-tubulin, neural stem cell markers such as Sox2 or prominin, and neurogenic determinant such as Ascl1 and Pax6. In this cellular context, retrovirus mediated co-expression of Ascl1 and Sox2 was sufficient to induce a full neuronal phenotype, whereas Ascl1 or Sox2 alone failed. The authors of the study hypothesised two possible scenarios to explain the synergism of the two TFs Ascl1 and Sox2 in the induction of the neuronal phenotype. They speculate that Sox2 might predispose the somatic genome of the pericytes to the neurogenic activity of Ascl1 or that the two TFs might interact in the regulation of common targets.

1.4.5 SoxC factors in neural development

1.4.5.1 Biological functions of the Sox C factors

Sox4, Sox11, and Sox12 constitute the group C of the family of Sox TFs. These genes have been highly conserved among vertebrate, and their biochemical similarities together with their expression pattern suggest that they act quite redundantly. Their expression has been found widespread and extensively overlapping in mouse embryo since mid-organogenesis. The highest RNA levels of these genes have been found in

post-mitotic neuronal progenitors throughout the nervous system, and also at lower level in mesenchymal cells of many developmental organs, (Penzo-Mendez, 2010).

As reviewed in Penzo-Mendez, it has been emerging that the SoxC genes have a critical role in the development of many organs, as consistent with their widespread expression. Sox4-null embryos and Sox11-null new-born die from major heart defects, and Sox11-null mutants display multiple defects in many other organs such as eyes, palate and lips, and lung. Sox4 is also required for the differentiation of B lymphocyte, pancreatic beta cells and osteoblast. Thus, the Sox factors are relevant for the differentiation of many cell lineages.

They also appear overexpressed in many types of cancer, and therefore they have been used as cancer markers, especially Sox11. These proteins are overexpressed in most medulloblastomas and gliomas, although their roles in cancer are not clear, as data reported are conflicting (Penzo-Mendez, 2010).

1.4.5.2 Molecular characteristics of the SoxC factors

The SoxC factors are single-exon genes, with the HMG DNA-binding domain in the N-terminal half of the protein and the transactivation domain (TAD) in the C-terminal side (Figure 1.7). Preferential binding of Sox4 HMG has been identified for more than one alternative motifs by EMSA experiments, as already mentioned in paragraph 1.4.1.2.1 and further reviewed in Penzo-Mandez (Penzo-Mendez, 2010). Interestingly, Sox4 binds more efficiently than Sox12 in EMSAs, and Sox11 binds very weakly (Dy et al., 2008). Different conformations of the TAD of the three SoxC proteins have been related to a different potent activation of their target, as proved by transient transfection reporter gene assays. In accordance with this observation, Sox11 is the most potent and

Sox12 the least potent of the three proteins in the activation of their targets (Dy et al., 2008). This is consistent with the fact that Sox12 is the less relevant and critical SoxC TF in the control of the differentiation processes in which these proteins have been implicated. It is also the factor showing the milder phenotype in knockout studies compared with the close related members of the same family.

SoxC proteins have been shown to cooperate with other TFs as partners. Intriguingly, like SoxB1/Sox2 factor(s), they strongly synergise with the POU domain TFs Brn1 and Brn2 in the activation of the *Fgf4* enhancer on the adjacent Sox-Pou binding motif (Dy et al., 2008). In the same way and as already previously mentioned (par 1.4.3), either Sox2 or Sox11 can synergistically activate *in vitro* the *Nestin* enhancer active in the early neural tube by binding the Sox-Pou motif there contained. *In vivo*, Sox2 and Sox11 are not coexpressed, but each of them is coexpressed with Brn2 and Nestin, further suggesting that the SoxC proteins and POU factors might synergise *in vivo* (Tanaka et al., 2004).

1.4.5.3 SoxC factors in neuronal differentiation

SoxC factors, Sox4 and Sox11, can also induce neuronal differentiation. A work by Bergsland and colleagues has shown that Sox4 and Sox11 have a critical role in the establishment of pan-neuronal proteins expression and neuronal traits (Bergsland et al., 2006). The authors have shown in their work that the ectopic expression of Sox11 (and equally Sox4), by *in ovo* electroporation, in the developing chick spinal cord can induce ectopic precocious expression of neuronal markers such as Tuj1 and Map2. Interestingly, the authors have proved that the SoxC factors act as transcriptional activators in the case of one of the neuronal proteins, more precisely by direct binding and activation of the promoter of the *Tubb3* gene. Sox4 and Sox11 motifs have been

identified in the promoter of *Tubb3* and the direct binding of the two SoxC TFs has been demonstrated in EMSA experiments. Also the transcriptional activation of the *Tubb3* promoter by the SoxC factors has been demonstrated by transient transfection and transcriptional reporter gene assay *in vitro*. Sox4 and Sox11 can initiate the expression of neuronal markers in cells that are still proliferating; therefore their control on the establishment of pan-neuronal properties in committed progenitor cells is independent from mechanisms that control cell cycle exit. Moreover, the authors have also shown that Sox4 and Sox11 are induced and act downstream of the proneural proteins, Ngn2 and Ascl1. In the same study, the transcriptional repressor protein REST/NRSF has been shown to control and restrict the domain of expression of the SoxC factors in post-mitotic differentiating neurons. Thus, the concomitant activity of proneural proteins and REST establish the pattern of expression of Sox4 and Sox11 in post-mitotic differentiating neurons in the developing chick spinal cord. In this way, the SoxC factors can control the program of neuronal differentiation and the establishment of neuronal traits downstream of the proneural proteins.

In analogy with their role in embryonic CNS, evidence shows that Sox4 and Sox11 induce neuronal differentiation also in the adult hippocampus in mouse. A work by Mu and colleagues has shown that SoxC factors are essential regulators of the genetic network controlling neuronal differentiation in adult neurogenesis (Mu et al., 2012). The expression of Sox4 and Sox11 is initiated at the time of neuronal fate commitment of hippocampal neural progenitors (in late Type2a and Type 2b cells) and is maintained in immature neurons. Retrovirus mediated overexpression of Sox4 and Sox11 promote *in vitro* neurogenesis from adult NSCs, as proved by the expression of neuronal markers such as DCX and *Tubb3* besides the proliferating-enhancing culture conditions of transduced cells compared with the control. Conversely, retrovirus Cre recombinase-

mediated deletion of Sox4 and Sox11 prevents *in vivo* and *in vitro* neurogenesis, as proved by depletion of the same neuronal markers in the Sox4-Sox11 double conditional knockout mice transduced with the Cre-retrovirus. Furthermore, in the same study the authors have shown that SoxC factors target the promoter of the neuronal differentiation gene DCX, as shown in ChIP experiments and luciferase assays. Therefore, another neuronal gene has been identified as direct target of Sox4 and Sox11, following the identification of Tubb3 by Bergsland et al., in the establishment of a neurogenic network of differentiation controlled by SoxC factors. Finally, it is worth noting that in this same study by Mu and colleagues, reprogramming of astroglia into neurons induced by Neurogenin2 is strongly enhanced by Sox4 and Sox11 factors. Surprisingly indeed, Neurog2-induced reprogramming is impaired in Sox4-Sox11 double conditional knockout astrocytes. These results show an essential role for Sox4 and Sox11 in neuronal reprogramming of astrocytes, providing further support for the view of SoxC factors as key regulators of the transcriptional network underlying neuronal differentiation.

1.4.6 Sequential binding of Sox factors in the specification of neural lineage

The overview of the literature on the Sox factors discussed in the previous paragraph has highlighted the notion that different classes of Sox factors are expressed during neurogenesis and act in different processes of the neural development from early lineage specification to neuronal differentiation.

A “priming” role has emerged for the Sox factors as regulators of the neurogenesis (Figure 1.8). This consists in the sequential occupation of common targets by different Sox factors at different stages of neurogenesis with the purpose of preselecting and

priming the bound genes for the activation at a later stage. Indeed, Bergsland and colleagues demonstrated in a genome-wide binding study that Sox2, Sox3 and Sox11 exhibit an notable sequential binding to a common set of neural genes (Bergsland et al., 2011). The binding profile of the three factors was exploited by ChIP-seq generating and comparing binding dataset for Sox2, Sox3 and Sox11 in mouse ESCs, ES cell-derived NSCs, and neurons as cellular sources (Figure 1.8). Interestingly, data has shown that in ESCs Sox2 binds and preselects for silent neural lineage-specific genes, which will be bound and activated later by Sox3 in NSCs. With an analogous pattern, in NSCs Sox3 binds and preselects genes that are later bound and activated by Sox11 in differentiating neurons. Genes preselected by the Sox proteins are in a poised state associated with a bivalent chromatin signature, which is then changed into a permissive chromatin state with the binding of the activating Sox factors. Thus, the entire neurogenesis can be controlled through an ordered succession of a group of Sox factors on a common set of target gene enhancers, a process to which Wegner referred as “Sox-session” in his commentary to the paper by Bergsland et al. published on the same issue of *Genes and Development* (Wegner, 2011). In conclusion, the paper of Bergsland and colleagues has proposed a pre-binding function for the Sox proteins involved in the pre-selection of targets that will be activated at a later stage of development. The priming factors might work by having an impact on the chromatin structure and epigenetic state of the preselected but not yet active target genes. This influence on the chromatin state has been proved at least for Sox3 in the conversion of a poised state of the neuronal enhancers still inactive in the NSCs.

Also in the specification of the B cells, Sox2 preselects the enhancer $\lambda 5$ -V preB1 that is later activated by Sox4, as described in a paper by Liber et al (Liber et al., 2010). Finally, Fox family of TFs has shown to operate a similar mechanism of regulation in

the activation of the liver-specific *Albumin* enhancer during development (Xu et al., 2009).

The new notion of the Sox factor family with members acting in succession in the establishment of the neurogenic program is a view that needs to be kept in consideration when investigating the molecular interactions of TFs and the Sox factors in the regulation of neural enhancers in development. Interactions between *Ascl1* and Sox factors are the object of investigation in this thesis.

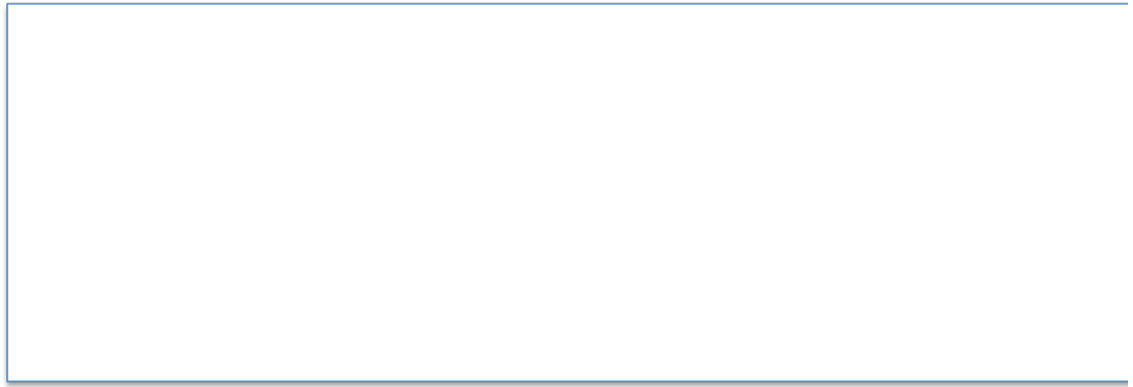


Figure 1.8 The Sox-session.

Model illustrating the sequential binding of Sox proteins to common downstream gene targets in the establishment of the neural lineage from ESCs to differentiating neurons. The cartoon also highlights the association between Sox pre-binding and bivalent histone modifications. Taken and adapted from *Bergsland et al., 2011*.

1.5 Use of NS5 cells as neural stem cell model system

The study of neural enhancers regulated by TFs *Ascl1* and *Sox* factors presented in this thesis has used an *in vitro* cell line with characteristics of neural stem cell as model system. The NS5 cells used in this study were first described in a work by Conti and colleagues (Conti et al., 2005). (The name NS5 refers to one of the first clones of the homogeneous adherent culture population for which the neural stem cell characteristics were analysed in details.)

Neural stem cells (NSCs) were first generated from mouse ES cells following neural induction upon serum withdrawal from the medium in adherent monolayer culture. The maintenance of NSC is strictly dependent on the presence of the growth factors EGF (epidermal growth factor) and FGF2 (fibroblast growth factor 2) in the culture medium. Upon the establishment of these culture conditions, *in vitro* NSCs independent from stem cell niche can be maintained and expanded as they undergo symmetric self-renewing divisions. They represent a pure homogeneous culture of neural stem cells. Even after prolonged expansion, they remain capable to differentiate into neurons and astrocytes *in vitro* with electrophysiological characteristics of functional neural cells. Withdrawal of FGF2 and EGF from the medium is sufficient to induce differentiation. The NS5 cells express morphological, cell biological and molecular features of radial glia cells. Indeed, NSCs can also be obtained from foetal mouse forebrain. Among the molecular features resembling radial glia, the neural stem cells express *Pax6*, *GLAST*, *BLBP* mRNA, and are immunopositive for nestin, RC2, vimentin, *Pax6* and prominin among radial glia markers. Furthermore, they express neural precursor markers such as *Sox2*, *Sox3*, *Emx2*, and the bHLH TFs *Olig2* and *Ascl1*. Indeed, coexpression of *Ascl1* and *Sox2* has been shown in NS5 cells by immunofluorescence staining for the purpose of the study presented in this thesis (Figure 1.9). While the expression of *Sox2* was

strong and uniform among all cells, *Ascl1* showed the typical “salt and pepper” expression pattern.

The expression of the TFs *Ascl1* and *Olig2*, together with a shown preference for differentiation into GABAergic neurons, exhibits a biased telencephalic ventral identity of the NS5 cells. These features might be the consequence of an *ex-vivo* environment and an effect of the exposure to FGF2 in culture. This aspect is suggestive of limitations generated by the *in vitro* environment and the impossibility to establish a connection between neural stem cells in culture and populations of radial glia progenitor cells *in vivo* in the CNS. Indeed, NS culture systems should be best regarded as an environment that force high rate of proliferation and repress regional or cell type specific differentiation. Moreover, the sustained exposure to growth factors in culture condition might alter the transcriptional and cellular phenotypes (Conti and Cattaneo, 2010). Alternative protocols for the derivation of radial glia *in vitro* in FGF and EGF-free culture conditions have been described with the generation of dorsal forebrain glutamatergic neurons. However, these cells are not easily expandable *in vitro* (Bibel et al., 2007, Bibel et al., 2004). Where other protocols have been considered, different drawbacks have emerged in the methods, as reviewed in Conti and Cattaneo, 2010 (Conti and Cattaneo, 2010).

Altogether, this suggests that beyond the limitations imposed from the culture conditions, the NS5 cells represent a good model of neural stem cells *in vitro*. They are high proliferating mitotic precursors, readily expandable cells, which divide nearly every 24 hours. Moreover, they can be transiently or stably transfected by electroporation, or lipofection, or viral transduction. In conclusion, they are versatile and easily manipulated (Conti et al., 2005). Therefore, they are well suitable to study

the regulation of gene expression by prominent TFs expressed in neural stem cells, which is the purpose of the study in this thesis.

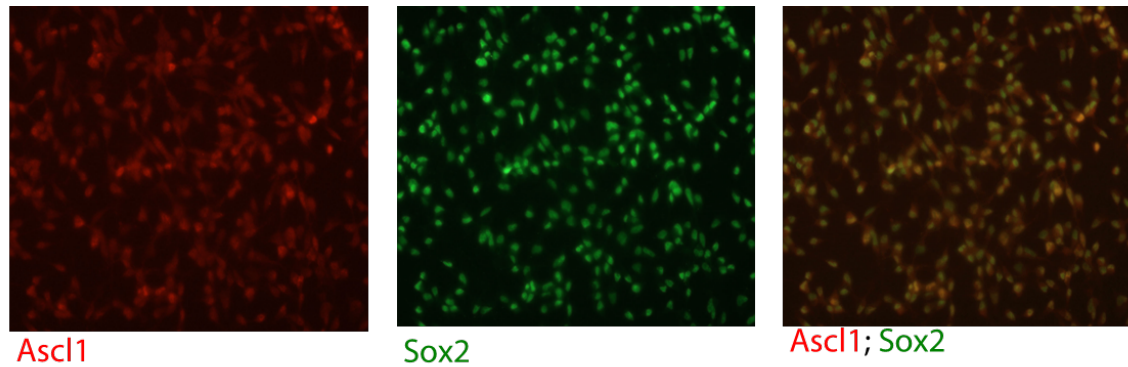


Figure 1.9 Coexpression of Ascl1 and Sox2 TFs in vitro in NS5 cells by immunofluorescence staining.

Sox2 shows a broad expression uniformly in all cells, whereas Ascl1 shows a typical “salt and pepper” expression pattern, as discussed elsewhere in chapter 1.3 Proneural proteins and the TF Ascl1 in vertebrate neural development. Images taken during this work.

1.6 Objectives of the thesis

This work aims at characterising transcriptional enhancers regulated by interplay between Ascl1 and Sox factors. It also aims to understand the mechanism of regulation by these TFs and the nature of their interactions.

Molecular data previously generated in our lab through a ChIP-seq approach identified genomic regions in NS5 cells, an *in vitro* model of neural stem cells, where Ascl1 and Sox2 bind in close proximity. In this work, a small group of nine regions selected from those identified by ChIP-seq were characterised as enhancers *in vitro* in NS5 cells through luciferase assays. This work also demonstrated that these enhancers are regulated by Ascl1 and Sox factors, in particular Sox2, SoxC, and Sox8. Luciferase assays of wt and mutant enhancers carrying disrupted binding motifs of Ascl1 and Sox factors were performed to identify and propose mechanisms of transcriptional regulation by these TFs through direct or indirect DNA binding to the identified motifs. Finally, luciferase assays were also performed after simultaneous overexpression of these TFs in NS5 cells to understand the nature of their interactions, for instance if they synergise or counteract each other in the regulation of the enhancers here identified.

2. Materials and Methods

2.1 Cell culture

2.1.1 NS5 cell culture

NS5 cells were maintained in NSC basal medium containing Euromed-N medium (Euroclone) supplemented with 1% L-glutamine (Gibco), 0.1% BSA (Gibco), 1% N2 supplement (100X) (R&D System), 1% penicillin and streptomycin (Pen/Strep) (Gibco), fibroblast growth factor-2 and epidermal growth factor (FGF2, EGF) (Peprotech), and laminin (Sigma), as described in Conti et al., (Conti et al., 2005). NS5 cells were cultured at 37°C in 5% CO₂ for 95% air humidified incubator.

2.1.2 Hek 293T cell culture

Hek 293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma) supplemented with 10% Foetal Calf Serum (Hyclone), 1% L-glutamine (Gibco), 1% penicillin and streptomycin (Pen/Strep) (Gibco). Cells were cultured at 37°C in 5% CO₂ for 95% air humidified incubator (Graham et al., 1977).

2.2 Chromatin immunoprecipitation (ChIP) and quantitative PCR (ChIP-qPCR)

2.2.1 Chromatin preparation

NS5 cells were fixed with Di-(N)-succinimydyl glutarate (DSG) and paraformaldehyde which establishes covalent links between proteins and DNA. Then, cells were lysed in

presence of SDS (1%). Chromatin was extracted with chloroform and sonicated with Bioruptor Sonicator (Diagenode), 200W, 30s ON/ 30s OFF, in an ice-cooled water bath for 15 min.

2.2.2 Chromatin immunoprecipitation (ChIP)

Chromatin input per immunoprecipitation (IP) was 50µg. The whole amount of chromatin needed for one IP was calculated depending on the number of conditions tested. A step to preclear the chromatin was performed, and chromatin was incubated for 2 hours with magnetic protein Dynabeads (Invitrogen) at 4°C. Magnetic beads were captured on a magnetic bar and the precleared chromatin was divided between IP tubes. A specific antibody was added and incubated overnight at 4°C. A commercial antibody (rabbit polyclonal p300 SantaCruz) to precipitate the protein p300 was used at 3µg/µl, after being tested at 3, 6 and 12 µg/µl. No antibody was added to the mock condition used as control. 50µg/µl of magnetic beads were added to capture immunocomplexes overnight. The protein beads have a strong affinity and specifically bind the antibodies allowing immunocomplexes to stay stacked on the magnetic bar between the washes. Before the washes, an input of chromatin was taken from the mock condition and diluted at 1%, 0.1% and 0.01% to establish the standard curves in the qPCR.

After the overnight IP chromatin was washed 3 times with reduced SDS buffer (50 mM HEPES-KOH, pH7.5; 150 mM NaCl; 2 mM EDTA; 1% TritonX-100; 0.1% NaDOC; 0.1% SDS), once with reduced buffer added with 0.35 M NaCl, twice with NP-40 wash buffer (10 mM Tris-HCl, pH8.0; 0.25 M LiCl; 1 mM EDTA; 0.5% NP-40; 0.5% NaDOC); once with TE solution. Then immunocomplexes were separated from protein beads with elution buffer (SDS 1%, Tris-HCl 50 mM, EDTA 10 mM) for 15 min at 65°C.

Then, cross-links were reversed to separate the proteins from DNA fragments by digesting with Proteinase K (100 µg/µl) in NaCl (200 µM) for 2 hours at 42°C and overnight at 65°C. Finally, the DNA fragments were extracted with phenol-chloroform, precipitated with glycogen and resuspended in water. (A similar protocol was followed for Ascl1 and Sox2 ChIP performed by Ben Martynoga, with specific antibody for the protein of interest used at the required concentration.)

2.2.3 Quantitative PCR (ChIP-qPCR)

Absolute qPCRs with specific forward and reverse primers listed in Table 2.1 were performed to amplify isolated DNA fragment from ChIP. SYBRGreen based kit for qPCR (iQ Supermix Biorad) was used for the amplification reactions. The same procedure described here for the p300 ChIP-qPCR was used also to amplify and quantify the targets in the Ascl1 and Sox2 IP chromatin from NS5 cells obtained from the ChIP-seq performed by Ben Martynoga. Standard curves were generated by serial dilutions of input DNA. IP efficiency was measured and expressed as fold enrichment, which was calculated as the fold change between the amounts of precipitated sequences for one target region divided by the amount of precipitated sequences of negative control regions (Delta1-ORF and/or Fbwx7-ORF).

2.3 Plasmids and cloning

2.3.1 Cloning the enhancer-regions into the luciferase reporter vector

The enhancer-regions tested in luciferase assay were amplified from mouse genome by PCR with high fidelity Pfu DNA polymerase (Stratagene). Specific primers flanking the regions of interest were designed, using the primer3 website (<http://frodo.wi.mit.edu/primer3/>). Primers are listed in Table 2.2. Restriction enzyme

sites for SalI and NheI (Roche) were added at the 5' end of the forward and reverse primers respectively, to digest and subclone the amplified inserts. The UCSC genome browser was used to determine the length of the enhancer-regions to clone. Usually the whole conserved chromatin regions among vertebrate species were chosen for cloning, likely to have a regulatory function to test, as described in the Results. The enhancer-regions amplified were cloned into a TOPO-Blunt vector (Invitrogen) using TOP10 competent cells (Invitrogen) and the Zero Blunt Topo PCR Cloning Kit according to the manufacturer's instructions. Subsequently, they were digested with SalI and NheI and subcloned into a p- β -Globin-Luciferase vector, also digested with SalI and NheI upstream of the β -Globin minimal promoter, ready to be tested in a luciferase assay.

2.3.2 Site-Directed Mutagenesis

Mutations to the E-boxes and Sox motifs of the enhancers were introduced using the QuikChange II XL Site-Directed Mutagenesis Kit from Stratagene according to manufacturer's instructions. Briefly, mutated primers listed in Table 2.3 carrying mutations in the sequence of the motifs were designed using the web software provided by Stratagene (<http://labtools.stratagene.com/QC>). The point mutations in the E-box and Sox motifs were designed as described in Castro et al., Tanaka et al., and in more details in Chapter 4 of the Results (Castro et al., 2006, Tanaka et al., 2004). Topo-vectors containing the wild-type enhancer-sequences were used as templates in the mutagenesis PCR. The resulting mutated enhancer-sequences were then subcloned into the p- β -Globin-Luciferase vector via digestion with restriction enzymes as described in the previous paragraph for the wild-type sequences. Mutated enhancers generated with this strategy were then tested for transcriptional activity in NS5 cells in luciferase assays performed as described in paragraph 2.4 for the wild-type constructs.

2.3.3 TFs expression vectors

The expression vectors listed in Table 2.5 were used to overexpress TFs of interest, in particular Ascl1 and the Sox factors, in NS5 cells and test the effect of their GoF on the activity of the enhancers in transcriptional assays. For the dominant repressor versions, the DNA binding domains of Ascl1 or Sox2 were fused with the *Drosophila* Engrailed repressor domain to generate pCAGGs-Ascl1-EnR or pCAG-Sox2-EnR vectors. These constructs were already available in the lab.

2.4 Luciferase assay

Prior to performing the transfection, NS5 cells were plated in 48 well plates containing NSC media. Around 150-200,000 cells were plated in each well. Luciferase constructs containing the enhancer-regions were tested in different conditions in triplicate or quadruplicate. Each condition contained the luciferase reporter vector (0.125 µg/µl) with or without the specific cloned enhancer-region, the pCMV-β-gal control vector (0.25 µg/µl) and either one or two TF expression vector(s) (0.25 µg/µl). Opti-MEM (Gibco) and Plus Reagent (Invitrogen) were used to complex the DNA plasmids over an incubation time of 15 minutes. The DNA complexes so formed were incubated for further 15 minutes and integrated into complexes with lipofectamine (Invitrogen). The complete NSC medium was removed from the wells and replaced by Opti-MEM medium to facilitate the transfection. The lipofectamine - DNA mix (0.625 µg/µl total DNA) was then added to the cells into the wells and incubated for 5 hours in 5% CO₂ at 37°C. After this first incubation, NSC medium replaced the Opti-MEM medium in the wells and the incubation was carried on overnight at 37°C.

After 24 hours the cells were finally lysed with Passive Lysis Buffer (Promega) and frozen at -80°C. A luminometer was used to measure the activity of the luciferase

enzyme after adding luciferine substrate (Promega) and the enzymatic activity of the β -gal by adding β -gal base solution supplemented with ONPG (Sigma) and DTT.

The luciferase activity was normalized to correct for variations in transfection efficiency for each condition between triplicates or quadruplicates, by dividing the luciferase activity for the β -gal signal obtained in corresponding wells. Data were then presented as the mean fold change relative to the negative control, that is the luciferase vector containing the enhancer region to test plus GFP expression vector only. The mean fold change of the basal activity of each regulatory region was also measured relatively to the activity of the empty vector.

2.5 Electrophoretic Mobility Shift Assay (EMSA)

2.5.1 Protein synthesis

2.5.1.1 In vitro transcription and translation

The full-length cDNA clones of the transcription factors listed in Table 2.6 were used as template for *in vitro* protein synthesis. Proteins were produced using the TnT[®] T7 Quick Coupled Transcription/Translation rabbit reticulocyte lysate system (Promega) in a 50 μ l reaction containing 40 μ l of TnT[®] rabbit reticulocyte lysate Master Mix and 1 μ g of DNA template, according to manufacturer's instructions. The reactions were incubated at 30°C for 90 minutes in a thermomix, according to manufacturer's instructions. Upon completion, reactions were snap frozen in liquid nitrogen and stored at -80°C until required.

2.5.1.2 Protein expression in 293T cell line

2.5.1.2.1 Transfection in 293T cells

293 cells were seeded in 10 cm Petri dishes in a number of 6 Millions in DMEM medium. After 24 hours, cells were transfected with pcDNA3.1-Ascl1-E47, pcDNA3.1-Sox2, or pCAG-Sox2 expression vectors. For transfection, 24 µg of plasmid DNA was mixed with Opti-MEM (Gibco) and Lipofectamine 2000 (Invitrogen) (pre-mixed and incubated for 5 minutes) and incubated for 20 minutes at room temperature. The DMEM medium was removed from the Petri dish and replaced with OPTI-MEM to facilitate the transfection. The lipofectamine - DNA mix was then added to the cells into the Petri dish and incubated for 5 hours in 5% CO₂ at 37°C. After this first incubation, fresh pre-warmed DMEM medium replaced the Opti-MEM in the dish and the incubation was carried on overnight at 37°C.

2.5.1.2.2 Cell lysis

After 24 hours, cells were washed and harvested in PBS with complete EDTA-free protease inhibitor (Roche), resuspended in lysis buffer (20 mM Hepes, pH 8.00; 100 mM KCl; 0.83 mM EDTA; 1.66 mM DTT; 1x complete EDTA-free protease inhibitor (Roche); 0.5% NP-40) (This buffer was modified from (Chew et al., 2005)) and rotated at 4°C for 30 minutes. After centrifugation at 13,000 rpm at 4°C for 15 minutes, the resulting supernatant was divided into 40 µl aliquots, which were snap frozen in liquid nitrogen and kept at -80°C until required. The total protein concentration in the lysate was quantified by performing Bradford Assay (Bio-rad) according to manufacturer's instructions (Bradford, 1976).

2.5.2 Radiolabelling of double stranded oligonucleotides

2.5.2.1 Annealing of the single stranded oligonucleotides

The single stranded oligonucleotides listed in Table 2.4 were resuspended in H₂O to a final concentration of 100 mM. Complementary oligonucleotide pairs were mixed in a 1:1 ratio (10 µg of each oligonucleotide) in TNE annealing buffer (10 mM Tris, pH 8.00; 50 mM NaCl; 1 mM EDTA) in a total volume of 100 µl. The mixture was heated to 95°C for 5 minutes in a hot block before being allowed to return to room temperature over a period of 2-3 hours. Annealed stocks were stored at -20°C.

2.5.2.2 Radiolabelling reaction

The 5' end of annealed oligonucleotides was radiolabelled by the addition of phosphorus 32 (³²P) using T4 polynucleotide kinase. A 20 µl reaction containing 200 ng of oligonucleotide, 1x T4 polynucleotide buffer (New England BioLabs), 10 U of T4 polynucleotide kinase (New England BioLabs), and 2 µl of ³²P-γATP (3000 Ci/mmol at 10 mCi/ml; PerkinElmer) was incubated at 37°C for 30 minutes. 80 µl of TNE buffer was added to the reaction before purification on an illustra microspin G-50 column (GE Healthcare) according to manufacturer's instructions. Radiolabelled oligonucleotides (at final dilution of 2 ng/µl) were stored at 4°C and used within one half-life (14 days).

2.5.3 Gel Shift Assay (EMSA)

Binding reaction included 2 µl of 5x binding buffer A, 1 µl of 1 µg/µl poly (dI-dC) (Thermo Scientific), 3 µl of *in vitro* translated protein (Ascl1-E47) or 2 µl of 293T cell lysate (overexpressing Ascl1-E47 or Sox2), and 1 µl of radiolabelled oligonucleotides (MSB4 probe) in a final volume of 10 µl. (The binding buffer A contains 20 mM Hepes,

pH 8.00; 100 mM KCl; 0.83 mM EDTA; 1.66 mM DTT; 1x complete EDTA-free Protease inhibitor (Roche); 20% glycerol). Binding reactions were incubated for 20 minutes at room temperature. Supershift assays were performed by adding 1 µg of primary antibody to the binding reaction with an incubation of further 20 minutes (mouse α -Ascl1 BD Biosciences) (goat α -Sox2 Santa Cruz sc-17320x). The addition of the antibody could be either before or after the addition of the radiolabelled probe to the binding reaction, as no differences were observed in the results in both cases. Competition assays were performed by the addition of a 200-fold excess non-labelled probe (cold probe) in the reaction. After incubations, the binding reaction mixtures were resolved on pre-run 5% non-denaturing polyacrylamide gel in 0.5x Tris-buffered EDTA (TBE). Electrophoresis was carried out at 200 V for over 2 hours in the cold room (4°C). The gel was dried under vacuum at 80 °C for 1 hour and exposed to X-ray film (Kodak) for 2 to 16 hours (depending on the strength of signal achieved). Films were developed for 2 minutes using a FPM-3800A processor (FujiFilm) and scanned with a Bio-rad GS800 densitometer (Bio-rad).

2.6 Western Blot

To assess protein synthesis either from using the *in vitro* TnT system or plasmid transfection in 293T cells, protein expression was analysed by SDS-PAGE protein separation and immunoblotting. Equal amounts of proteins from samples in SDS gel loading buffer were loaded onto 4% to 12% gradient polyacrylamide gel (NuPAGE bis-Tris gels) (Invitrogen) and separated by electrophoresis according to their molecular weight, under reducing conditions. Proteins were then transferred onto nitrocellulose membranes (GE Healthcare) under a current of 250mA in transfer buffer (transfer buffer per litre: 3 g trizma base, 14.4 g glycine, 150 ml methanol). Membranes were blocked in Tris-Buffered Saline-Tween (TBS-T) with 10% milk (TBS-T per litre: 1.21

g trizma base, 11.68 g NaCl, 0.5 ml tween, pH to 7.4 with HCl) for 30 minutes at room temperature and then incubated overnight with the primary antibody diluted in TBS-T with milk at 4°C (Primary antibody: mouse α -Ascl1 1:200 Hybridoma Bank; goat α -Sox2 Santa Cruz 17320x 1:500). Membranes were washed with TBS-T, incubated with the appropriate secondary antibody diluted in TBS-T for 2 hours at room temperature (Secondary antibody: HRP-conjugated anti-mouse or HRP-conjugated anti-goat 1:10.000, DAKO). After further washes, chemiluminescent detection was performed using Amersham ECL (GE Healthcare) according to manufacturer's instructions. Films were developed for 2 minutes using a FPM-3800A processor (FujiFilm) and scanned with a Bio-rad GS800 densitometer (Bio-rad).

2.7 Computational analysis

UCSC genome browser and PhastCons program were used to assess the sequence conservation of the genomic regions bound by Ascl1 and Sox2 (identified by ChIP-seq) among vertebrate species (<http://genome.ucsc.edu/>) (Siepel et al., 2005). The search for the Ascl1 and Sox2 binding motifs consensus was performed using the TESS (Transcription Element Search System) database (<http://www.cbil.upenn.edu/cgi-bin/tess/tess>), searching for consensus position weight matrices from the TRANSFAC database (Wingender et al., 1996). The expression pattern of the genes associated to Ascl1 and Sox2 bound regions was screened using Genepaint (<http://www.genepaint.org/>) database.

2.8 Statistical analysis

One-way ANOVA followed by Dunnett's post-test was used for the analysis of data from the basal luciferase activity of wt or mutant enhancers (where the activity of each enhancer was compared to the activity of the empty vector used as control) (Figures 3.7 and 4.2). Two-way ANOVA was employed for the analysis of the luciferase activity of enhancers (wt or mutant enhancers) following overexpression of TFs. In cases where all columns are compared to each other (where various combinations of TFs were overexpressed), a Tukey's post-test was used (Figures 3.8, 3.11, and 4.4). Where each column was compared to its control column (for example, where the activity of mutant enhancers following overexpression of a given TF were compared to the activity of the wt enhancer following overexpression of that same TF), a Dunnett's post-test was used (Figures 4.3, 4.5, 4.6, and 4.7). Differences were considered statistically significant at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$. All data are presented as the mean \pm SD. Values were calculated using GraphPad Prism 6.0c.

2.9 Materials

Oligonucleotides and Plasmids

Table 2.1 ChIP-qPCR primers

Primers to amplify ChIP and ChIP-seq identified regions bound by Ascl1, Sox2, and p300 in IP chromatin from NS5 cells

Primers	Primer Sequence (5'-3')
MSB 4 For	AGGTAGACAGGACGGAAAGG
MSB 4 Rev	TTTCACACCCAACAATGGAT
MSB 9 For	CAGACACTACTCCAGCAGAGG
MSB 9 Rev	AGACTAAAGGGAAGGCAGGA
MSB 10 For	GAGTGTGCCCATTGTTCTTT
MSB 10 Rev	GTGCCAAGATAGGTGACAGG
MSB 11 For	AGGGACTCACAGACACCAGA
MSB 11 Rev	ATAAAACAGCCATCGGACCT
MSB 18 For	GGTGAGGGGAAGTAGGAAGA
MSB 18 Rev	ATGAGAAGGCGCTGTTTAGA
MSB 22 For	ATGGGCAGAGTGTCATCAAT
MSB 22 Rev	ACCTGACTTCCTTCCTCCAG
MSB 23 For	GCATCTAAAGGGCCAGTATG
MSB 23 Rev	TGGTTCCTGGTTTGA CTCTG

MSB 24 For	GGGTCCTGCACAATAGCTC
MSB 24 Rev	CCTTACCTCCAAAACCCAAG
MSB 25 For	TTGGGGTAAAGCTACATTGC
MSB 25 Rev	GCCTGTGTTTGTAGTCCCTCCT
MSB 26 For	GCTGCTACTAAAGCCAAGGTT
MSB 26 Rev	CATTACAGCACAGCCATC
NB 1 For	CCAATAAAAAGGCTGTGGAA
NB 1 Rev	ACACAAATCTGGGGTTGGTA
NB 2 For	ACTTTGTAGGCAGAGGCAGA
NB 2 Rev	CCAGTGCTGAGATCAAGACA
NB 3 For	GGCCTAGATCGAATGTGAAG
NB 3 Rev	GGGCAGTAAAAGCAGAAACA
NB 4 For	CTCAGGCAGGTGTAGTCTGG
NB 4 Rev	CTCTTAGCAATGGTCACTTGG

Table 2.2 Cloning primers

Primers for cloning enhancer-regions in p- β -Globin-luciferase vector

Primers	Primer Sequence (5'-3')
MSB 4 For	CATGTAGTCGACAGGACTCACTGGGAACCTCT
MSB 4 Rev	GATCTAGCTAGCCATCAGGGGGATGGATCT
MSB 10 For	CATGTAGTCGACCAGTGTGGTTTATGGTTTTGG

MSB 10 Rev	GATCTAGCTAGCCAGAAAATTAAAGTGCGACCTC
MSB 11 For	CATGTAGTCGACAAAGAACTCGGAAGTGGACA
MSB 11 Rev	GATCTAGCTAGCCCATTTGGTCCTGTGAGATG
MSB 18 For	CATGTAGTCGACGCCAGGGCTTTACAGAGAA
MSB 18 Rev	GATCTAGCTAGCGGGTTGGAGTGAGTTTTTCC
MSB 22 For	CATGTAGTCGACCCCCTCCCTACACACATACA
MSB 22 Rev	GATCTAGCTAGCGAGGAAAAACACCACACCTG
MSB22-short For	GTCGACAGGGAGTGTGTGCTGGAAT
MSB22-short Rev	GCTAGCGAGCCTATGTCAATGGGTACA
MSB 23 For	CATGTAGTCGACCTGACCCAGACAGGACAAAC
MSB 23 Rev	GATCTAGCTAGCAGGGACCTTGGAGTCGTAAT
MSB 24 For	CATGTAGTCGACGGTCTCCATTTCTCCTTTGC
MSB 24 Rev	GATCTAGCTAGCTTGACCCCCTGACTGATTTA
MSB24-short For	GTCGACCCCAGAGGAAGAGAAAAAGG
MSB24-short Rev	GCTAGCTTGACCCCCTGACTGATTTA
MSB 25 For	CATGTAGTCGACCATCTGTTTTGCCTTTGTGTC
MSB 25 Rev	GATCTAGCTAGCCCATGCTAAATTCCTAGTGAGC
MSB 26 For	CATGTAGTCGACCGGGAAGAAGAAAGTGGAAC
MSB 26 Rev	GATCTAGCTAGCTGATTAGGAATGCTCCAGAAAC

Table 2.3 Primers for site-directed mutagenesis

Primers for site directed mutagenesis of the TF binding motif sequences to generate mutant enhancers

Primers	Primer Sequence (5'-3')
MSB4-M1 For	GCTCTCTCTCCAGTGTAGCCTCCAGGCCCTTGCTGC
MSB4-M1 Rev	GCAGCAAGGGCCTGGAGGCTACACTGGAGAGAGAGC
MSB4-M2 For	CTCTGAGTCCCCTGGGGCTCATCCAGCACTGGGTGTGAAAGCAG
MSB4-M2 Rev	CTGCTTTCACACCCAGTGCTGGATGAGCCCCAGGGGACTCAGAG
MSB11-M1 For	CCGGACTGTCCACATCTAGCCTGAGGAGCCAACACTGT
MSB11-M1 Rev	ACAGTGTTGGCTCCTCAGGCTAGATGTGGACAGTCCGG
MSB11-M2 For	GAGCCCCAAAGAACAAAGCCAATAGCCTTCTTTTCAGGTCCGATG
MSB11-M2 Rev	CATCGGACCTGAAAAGAAGGCTATTGGCTTTGTCTTTGGGGCTC
MSB11-M3 For	CTGTTCTTGTGTGTCTGCTTGAGCCCCAAAGAGTGCTGCCAACAGCTGTCTTTT
MSB11-M3 Rev	AAAAGACAGCTGTTGGCAGCACTCTTTGGGGCTCAAGCAGAACACACAAGAACAG
MSB11-M5 For	CTGTTCTTGTGTGTCTGCTTGAGCCCCAAAGAGTGCTGCCAATAGCCTTCTTTT
MSB11-M5 Rev	AAAAGAAGGCTATTGGCAGCACTCTTTGGGGCTCAAGCAGAACACACAAGAACAG
MSB18-M1 For	CTTTCATGCCCCCAAGTAGCCTGTGCCACTCCAGACTC
MSB18-M1 Rev	GAGTCTGGAGTGGCACAGGCTACTTGGGGGCATGAAAG
MSB18-M2 For	TGCCTGCCTTGTGCTAGCCTCTCCAGCTGCCCT
MSB18-M2 Rev	AGGGCAGCTGGGAGAGGCTAGCACAAAGGCAGGCA

MSB18-M3 For	GTGCCAGCTGCTCCTAGCCTCCCTGGCTGCAGGC
MSB18-M3 Rev	GCCTGCAGCCAGGGAGGCTAGGAGCAGCTGGCAC
MSB18-M4 For	GAAGGACCAGGCCAGAACTGATTGTGCTGTCCCTCCAGGGGCCCC
MSB18-M4 Rev	GGGGCCCCCTGGAGGGACAGCACAAATCAGTTTCTGGCCTGGTCCTTC
MSB18-M5 For	CAGCTGGTGCCACTCCAGACTCGTGCTGGGAGGCTCATTCTAAACAGCG
MSB18-M5 Rev	CGCTGTTTAGAATGAGCCTCCCAGCACGAGTCTGGAGTGGCACCAGCTG
MSB22s-M1 For	GAGTGTCATCAATGTGGTTGTAGCCTCAGGGACTGTTCTCTGTGTG
MSB22s-M1 Rev	CACACAGAGAACAGTCCCTGAGGCTACAACCACATTGATGACACTC
MSB22s-M2 For	TCTGTGTGGTCCAAGTCACTCACTAGTGCTGCTGTCTTCTTGTATCCAGTC
MSB22s-M2 Rev	GACTGGATACAAGAAGACAGCAGCACTAGTGAGTGACTTGGACCACACAGA
MSB22s-M3 For	GTGTTTGTTCAGCATCTCTTGGCGTGCTGGGCAGAGTGCATCAATGTGG
MSB22s-M3 Rev	CCACATTGATGACACTCTGCCCAGCACGCCAAGAGATGCTGAACAAACAC
MSB24s-M1 For	CTTTCAGCCTCTGCAGTAGGCTTTGGGAGGGCTCTGGA
MSB24s-M1 Rev	TCCAGAGCCCTCCCAAAGCCTACTGCAGAGGCTGAAAAG
MSB24s-M2 For	GATGCAAAATCTGGACTGGATTAAAGTCCACTAGTGCAGGATTGAGTCTCTGCTTG
MSB24s-M2 Rev	CAAGCAGAGACTCAATCCTGCACTAGTGGACTTAAATCCAGTCCAGATTTTGCATC

Table 2.4 DNA Oligonucleotides probes for EMSA experiments

Probe	Sequence (5'-3')
MSB4-wt For	ccagtgcagctgccagggcccttgctgcctctgagtcacctggggctcatccattgtgggtgt
MSB4-wt RevCom	acaccaacaatggatgagccccaggggactcagaggcagcaagggcctggcagctgcactgg

MSB4-M1 For	ccagtgtagcctccaggcccttgctgcctctgagtcacctggggctcatccattgtgggtgt
MSB4-M1 RevCom	acaccaacaatggatgagccccaggggactcagaggcagcaaggcctggaggctacactgg
MSB4-M2 For	ccagtgcagctgccaggcccttgctgcctctgagtcacctggggctcatccagcactgggtgt
MSB4-M2 RevCom	acaccagtgctggatgagccccaggggactcagaggcagcaaggcctggcagctgcactgg

Table 2.5 Expression vectors used in transfection and luciferase assays

Plasmid	Full-length cDNA	Sources
pCAGGs-Ascl1	rat Ascl1	Guillemot's lab
pCAG-Sox2	mouse Sox2	Guillemot's lab
pCMV-Sox11	mouse Sox11	Guillemot's lab
pcDNA3.1-Sox4	mouse Sox4	Lovell-Badge's lab
pCAGGs-Sox8	chick Sox8	Briscoe's lab
pCAGGS-Ascl1-EnR [†]		Guillemot's lab
pCAG-Sox2-EnR [†]		Guillemot's lab
p-β-Globin-luciferase*		Guillemot's lab
pCMV-β-Gal		Guillemot's lab

[†] The cDNA sequence of Ascl1 or Sox2 is fused with the Drosophila Engrailed repressor domain.

* All enhancer-containing-β-Globin-luciferase vectors used in the luciferase assays were generated as described in paragraph 2.3.1 and 2.3.2 for mutant enhancers.

Table 2.6 Expression vectors for protein synthesis for EMSA assays

Plasmid	Full-length cDNA	Sources
pcDNA3.1-Ascl1-E47	rat Ascl1 tethered to mouse E47	Guillemot's lab
pcDNA3.1-Sox2	mouse Sox2	Cloned during this work
pcDNA3.1-Sox4	mouse Sox4	Lovell-Badge's lab
pcDNA3.1-Sox11	mouse Sox11	Cloned during this work

3. Genome-wide identification of neural enhancers regulated by Ascl1 and Sox2

3.1 Introduction

The work of this thesis aims to investigate interactions between the TFs (Transcription Factors) Ascl1/Mash1 and several Sox factors, primarily Sox2 as transcriptional regulators of neural enhancers in neurogenesis. Molecular data obtained in our lab through a genome-wide ChIP-seq approach (Chromatin Immunoprecipitations followed by high throughput sequencing) identified genomic elements in NS5 cells where Ascl1 and Sox2 bind in close proximity. We first wanted to validate the ChIP-seq results and identify and characterise a discrete group of regulatory elements bound and regulated by Ascl1 and Sox2. For this purpose, we selected a group of genomic regions commonly bound by Ascl1 and Sox2 with ideal characteristics such as association to genes showing deregulation by the perturbation of the two TFs in microarray experiments. Following this, to strengthen the possibility that the bound regions identified act as regulatory elements, I analysed the expression pattern in the CNS, the telencephalon in particular, of the genes associated to the bound regions and the conservation of the genomic sequences among vertebrate species. Finally, to test the transcriptional activity of the identified genomic elements and dissect whether they could act as enhancers or repressive elements in neural stem cells I performed luciferase assays. In this chapter, selection of the bound regions and their features and characteristics are discussed. Results to prove their transcriptional potential are presented.

3.2 ChIP-seq identification of genomic regions bound by Ascl1 and Sox2: computational analysis (Ben Martynoga's data)

The genomic elements characterized in this study were identified through a ChIP-seq approach. The first step towards the identification of genomic regions bound and regulated by Ascl1 and Sox2 is the validation of the data from ChIP-seq experiments, and the selection of elements with putative regulatory functions. ChIP-seq experiments were performed in the lab on NS5 cells to identify genome-wide regions bound by Ascl1 and Sox2. Dr. Ben Martynoga, a post-doc in our lab, performed the ChIP-seq experiments. This paragraph discusses the computational analysis that Ben Martynoga carried out in order to identify regions bound by both TFs.

NS5 cells have characteristics of neural stem cells and endogenously express the TFs Ascl1 and Sox2 (Conti et al., 2005, Castro et al., 2011) (Figure 1.9 in Introduction). Antibodies against Ascl1 and Sox2 were used in two independent ChIP-seq experiments to obtain an Ascl1 immunoprecipitated chromatin and a Sox2 immunoprecipitated chromatin. Using this method, in an early ChIP-seq dataset 3154 bound regions were identified for Ascl1 and 3838 bound regions for Sox2. Aiming to identify genomic regions commonly bound by Ascl1 and Sox2, Ben Martynoga identified 281 elements where Ascl1 and Sox2 peaks (regions of significant binding defined by the MACS peak-caller - Model-based Analysis of ChIP-Seq data) overlapped by at least 1bp, suggesting that these two TFs might bind in close proximity on these regions (Figure 3.1) (Zhang et al., 2008). Next, the bound regions were associated with their closest gene, upstream or downstream, since the purpose of the analysis was the identification genome-wide of the regions bound by the two TFs and their regulated genes. The assumption made is that the closest gene to the TF binding site is the most likely one to

be regulated by it. According to this, a list of 281 genes associated to genomic regions potentially bound by Ascl1 and Sox2 was generated. To further focus on genes likely to be directly regulated by Ascl1 and Sox2, the list of 281 genes associated to the bound regions was compared with 4304 genes showing de-regulation following perturbation of Ascl1 or Sox2 in microarray experiments in NS5 cells (previously performed in our lab). In these experiments either Ascl1 or Sox2, and their dominant repressive versions, Ascl1EnR or Sox2EnR, were overexpressed in NS5 cells and the effect on gene expression was analyzed. Ascl1-EnR and Sox2-EnR vectors consist of only the DNA binding domain of the TFs fused to the *Drosophila* Engrailed repressor (EnR) domain, which is able to repress the transcriptional activity of a bound region. As a result of the intersection of the ChIP-seq data and the microarray data, 40 genomic regions were finally identified as potentially bound and regulated by Ascl1 and Sox2 (Figure 3.1). About their genomic location, 50% of the 40 regions were intergenic and 50% were located within introns (data not shown). Also, in term of their distance from the transcriptional start site (TSS) of the closest gene, 30% of the regions were located at a distance farther than 100 Kb from the TSS (>100 Kb), 37.5% were between 10 and 100 Kb, and finally 32.5% of them were located within 10 Kb from the TSS (<10 Kb). The 40 regions of this early ChIP-seq dataset and their characteristics, such as their genomic coordinates, name of the closest gene associated to them, and distance from TSS, are listed in Table 3.1. Each genomic region has been given an arbitrary name or region ID as shown in the table. Hereafter in this thesis, the bound regions identified by ChIP-seq are named after their ID names.

In conclusion, Ben Martynoga obtained from the computational analysis a list of 40 regions bound and regulated by Ascl1 and Sox2 in an early ChIP-seq dataset available

when I started my project (Figure 3.1). A final dataset has been obtained by Ben Martynoga later during my project, as explained in paragraph 3.3.2

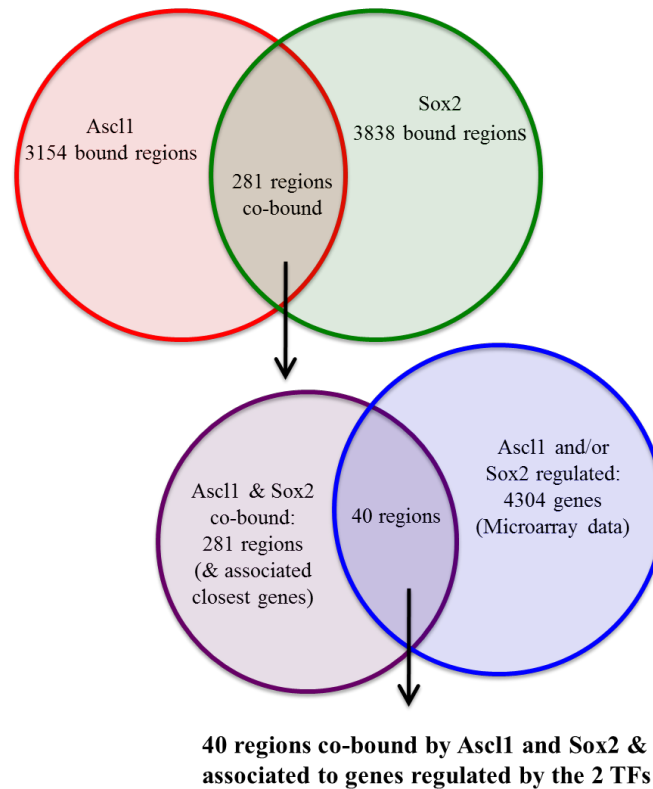


Figure 3.1. Computational analysis of Ascl1 and Sox2 ChIP-seq data.

The Venn diagrams show the computational analysis of the ChIP-seq data to identify candidate regulatory regions co-bound and regulated by Ascl1 and Sox2. Regions bound by Ascl1 and Sox2 were identified in two independent ChIP-seq experiments. 281 regions were identified as co-bound by both TFs (regions where Ascl1 and Sox2 peaks overlapped by at least 1 bp). The 281 co-bound regions were associated to the closest genes. The list of 281 genes was compared with a list of 4304 genes de-regulated by perturbation of Ascl1 and/or Sox2 obtained by microarray experiments previously performed in the lab. The purpose was to focus on genes likely to be regulated by the two TFs. As a result of the intersection of ChIP-seq and microarray data, 40 regions were identified as potentially bound and regulated by Ascl1 and Sox2. From *Ben Martynoga, unpublished data*.

Region-ID name	Genomic coordinates			Closest gene Name	TSS Distance
	Chrom	Start	End		
MSB 1	chr1	43751791	43751987	1500015010F	35706
MSB 2	chr1	90555009	90555220	Arl4c	43231
MSB 3	chr4	108897929	108898117	Calr4	8973
MSB 4	chr15	86018783	86018975	Cerk	-2629
MSB 5	chr9	110108555	110108770	Cspg5	37517
MSB 6	chr3	105388707	105388880	Ddx20	101484
MSB 7	chr8	83975980	83976141	Inpp4b	262958
MSB 8	chr7	25158765	25158907	Kcnn4	-3432
MSB 9	chr3	88298907	88299090	Lmna	8139
MSB 10	chr1	108098606	108098821	Phlpp	-30025
MSB 11	chr19	38714386	38714558	Plce1	-115699
MSB 12	chr2	151429394	151429604	Snph	28699
MSB 13	chr10	105076011	105076180	Tmtc2	-64476
MSB 14	chr13	60027356	60027620	Zcchc6	-102894
MSB 15	chr17	31694480	31694680	1500032D16F	-37460
MSB 16	chr7	141419585	141419785	Adam12	-2905
MSB 17	chr4	40883432	40884162	Bag1	11517
MSB 18	chr9	62379553	62379888	Coro2b	5131
MSB 19	chr10	98594851	98595142	Dusp6	130869
MSB 20	chr11	61342702	61342902	Epn2	50385
MSB 21	chr15	90908365	90908789	Kif21a	-28308
MSB 22	chr10	7325073	7325273	Lrp11	-15514
MSB 23	chr17	31694480	31694680	Pknox1	7166
MSB 24	chr1	196573235	196573435	Plxna2	-127312
MSB 25	chr18	47962648	47962848	Sema6a	-434226
MSB 26	chr9	110108613	110108813	Smarcc1	-74172
MSB 27	chr12	28260592	28260792	Sox11	-233253
NB 1	chr10	21862105	21862294	E030030I06R	6335
NB 2	chr10	21862440	21862714	E030030I06R	5915
NB 3	chr10	21862739	21862831	E030030I06R	5798
NB 4	chr6	94776827	94776936	Lrig1	-126675
NB 5	chr2	124618500	124618700	Sema6d	-702895
NB 6	chr9	26327927	26328133	B3gat1	213303
NB 7	chr13	97960486	97960612	Hexb	7636
NB 8	chr18	40467704	40468028	Kctd16	-49689
NB 9	chr12	68160131	68160961	Mdga2	161275
NB 10	chr5	115372437	115372653	Oasl1	596
NB 11	chr18	13107708	13107852	Osbpl1a	-7422
NB 12	chr18	43135589	43135693	Ppp2r2b	83201
NB 13	chr9	3258803	3259015	Alkbh8	76569

Table 3.1. Genomic regions co-bound by Ascl1 and Sox2.

The table shows the 40 regions co-bound and regulated by Ascl1 and Sox2 identified through computational analysis of ChIP-seq data. For each bound region, genomic coordinates, name of the closest gene associated to the region, and distance from the Transcriptional Start Site (TSS) are shown in the table. An arbitrary name was given to each region. This is shown in the first column of the table as Region-ID name. Hereafter in the thesis, bound regions/regulatory elements are named after their region IDs.

3.3 Selection of putative regulatory elements from *Ascl1* and *Sox2* co-bound regions identified by ChIP-seq

Next, we aimed to identify candidate regulatory elements bound and regulated by *Ascl1* and *Sox2* among the 40 regions identified by ChIP-seq. We randomly selected a discrete number of regions, in order to limit the number of elements to validate functionally in further experiments, as the initial number of 40 was too high. Moreover, for the regions randomly selected, we set out to determine whether they have properties of *bona fide* enhancers in terms of expression pattern of the associated genes, sequence conservation, and chromatin signature. These characteristics of the selected regions are described over the next four paragraphs.

3.3.1 Expression pattern in the telencephalon of the genes associated to the bound regions

For the bound regions selected, we considered the expression pattern of their associated genes. We focused on genes that are expressed in the mouse developing CNS, particularly in the telencephalon, since their expression pattern can be consistent with a regulation by *Ascl1* and *Sox2*, also expressed in the telencephalon. Moreover, their expression can be indicative of a function in the development of the CNS. We carried out an analysis of the expression patterns of these genes using publicly available *in situ* hybridization (ISH) data from the database Genepaint (<http://www.genepaint.org/>).

The following 14 bound regions were chosen: MSB4, MSB9, MSB10, MSB11, MSB18, MSB22, MSB23, MSB24, MSB25, MSB26, NB1, NB2, NB3, and NB4. They are all associated to genes that are expressed in the telencephalon, as shown in Figure 3.2. Therefore, they represent a good dataset of *bona fide* candidates for regulation by

Ascl1 and Sox2 (NB: The 3 peaks NB1, NB2 and NB3 are associated to the same closest gene: E030030I06Rik). From the 12 genes, 10 are expressed in the ventricular zone of the mouse telencephalon at embryonic stage E14.5, in the same expression domain of Ascl1 and Sox2. The other 2 genes, Lrp11 and Plxna2 (associated to peaks MSB22 and MSB24 respectively), show expression in the mantle zone (MZ) of the mouse telencephalon at the same stage E14.5 (Figure 3.2).

We also analysed the sequence conservation of the selected bound regions among vertebrate species. This analysis is described in paragraph 3.3.3 since, for each element selected, it was extended to the entire larger putative regulatory region cloned into the reporter vector and tested for transcriptional assays rather than only to the ChIP-seq bound region identified, as better explained later.



Figure 3.2. Expression pattern of the genes associated to the candidate bound regions.

Mouse telencephalon sections, E14.5, ISH, from Genepaint. For each gene, gene name/region ID name is labeled at the bottom of the picture. Lrp11 and Plxna2 are expressed in Mantle Zone (MZ). All other genes are expressed in the Ventricular Zone (VZ). Expression pattern of Ascl1 and Sox2 at the same stage is also shown.

3.3.2 Validation of the selected Ascl1 and Sox2 co-bound regions by ChIP-qPCR

To ensure the reliability of the ChIP-seq data and confirm recruitment of Ascl1 and Sox2, we performed ChIP-qPCR on the 14 bound regions selected as putative regulatory elements.

Primers flanking the bound regions (Table 2.1 in Materials and Methods) were used to amplify these elements in chromatin from NS5 cells immunoprecipitated either with an α -Ascl1 antibody or with an α -Sox2 antibody compared with a mock immunoprecipitated chromatin as control. Delta1 and Fbw7 coding sequences (Dll1 ORF and Fbw7 ORF) not precipitated by Ascl1 and Sox2 antibodies were used as negative controls (Figure 3.3). Enrichment of each region was given as fold change over the enrichment of negative controls. A threshold value of 2 was set, so that regions with enrichment higher than 2 were considered as positive and bound by Ascl1 and Sox2. For each region, values are shown in mock, α -Ascl1 and α -Sox2 immunoprecipitated chromatins for comparison. Hom2 and Sox21 represented positive controls as enhancers known to be regulated by Ascl1 and Sox2, respectively (Castro et al., 2006). Indeed, they were highly enriched in the respective α -Ascl1 and α -Sox2 immunoprecipitated chromatins, as shown in Figure 3.3.

The following nine regions, MSB4, MSB10, MSB11, MSB18, MSB22, MSB23, MSB24, MSB25, and MSB26, were strongly enriched in both α -Ascl1 and α -Sox2 immunoprecipitated chromatins, with values higher than 2, and therefore considered bound by Ascl1 and Sox2. Region MSB9 instead showed enrichment in α -Ascl1 immunoprecipitated chromatin, but had a value lower than 2 in α -Sox2 immunoprecipitated chromatin, therefore was excluded from our group of candidates, as

we were looking for genomic regions commonly bound by the two TFs. Finally, regions NB1, NB2, NB3, and NB4 were not precipitated compared to the negative controls and were therefore excluded from our group of candidates, as not bound by the two TFs.

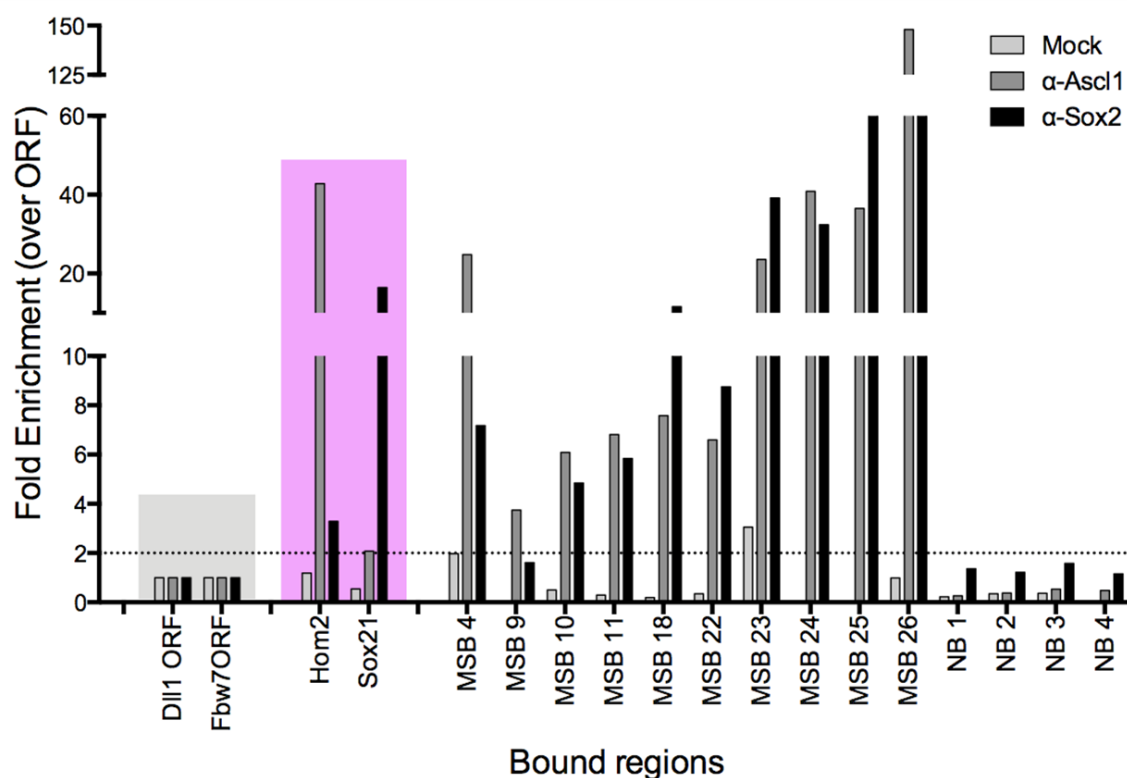


Figure 3.3. Validation of selected regions bound by Ascl1 and Sox2 by ChIP-qPCR.

Chromatin from NS5 cells immunoprecipitated with α -Ascl1 and α -Sox2 antibody compared to the mock immunoprecipitated chromatin. Negative controls and positive controls on grey and violet background respectively. A total of 14 regions were tested, of which 10 were validated as bound by the 2 TFs, with fold enrichment values higher than 2, set as threshold (fold enrichment over the negative controls). Regions NB1, NB2, NB3, and NB4 were validated as not bound by Ascl1 and Sox2.

These results of the ChIP-qPCR are consistent with final ChIP-seq data generated from Ben Martynoga later during my project (Figure 3.4). In his final dataset, Ben identified 2910 regions co-bound by Ascl1 and Sox2. This result suggests that Ascl1 and Sox2 co-binding occurs quite commonly in the neural stem cell genome. When the 40 original

candidate regions from the earlier dataset were compared and overlapped with the final data, 27 of them were validated and confirmed as co-bound regions in the final ChIP-seq dataset. The 9 regions I confirmed as bound by Ascl1 and Sox2 by ChIP-qPCR are included in the 27. However, 13 regions from the early 40 candidates were not validated in the final ChIP-seq dataset. These 13 include the 4 regions I could not confirm as bound in the ChIP-qPCR experiment.

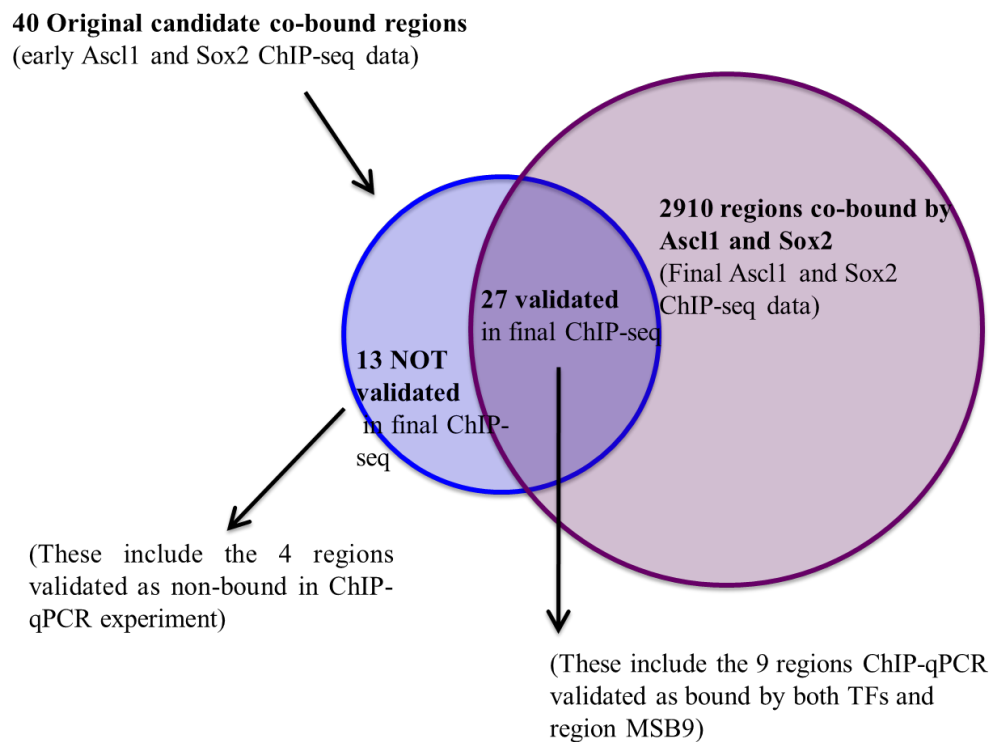


Figure 3.4. Comparison of early and final ChIP-seq data and ChIP-qPCR validation of regions bound by Ascl1 and Sox2.

27 of the 40 candidate regions co-bound by Ascl1 and Sox2 from early ChIP-seq data have been validated also in the final ChIP-seq data. These include the 9 regions validated as bound by Ascl1 and Sox2 in ChIP-qPCR. 13 regions of the 40 co-bound from early data have not been validated in the final ChIP-seq data. These include the 4 regions that were considered not bound by Ascl1 and Sox2 according to ChIP-qPCR results and therefore excluded from the dataset of the candidate enhancers.

In conclusion, the ChIP-qPCR confirmed recruitment of both Ascl1 and Sox2 on 9 out of 10 regions. These genomic regions represent our dataset of putative regulatory elements bound by Ascl1 and Sox2. Our next aim was to functionally evaluate their transcriptional activity and their regulation by Ascl1 and Sox2.

3.3.3 Sequence conservation of the genomic elements

We looked at the sequence conservation of the bound regions among vertebrate species. Comparative genomics has successfully identified transcriptional enhancers among evolutionary conserved non-coding sequences in the vertebrate genome (Pennacchio et al., 2006, Visel et al., 2009b). Indeed, functional regulatory sequences tend to be evolutionary constrained, so that non-coding regions conserved and constrained among the genomes of different species are more likely to have a function as non-coding regulatory elements. With this notion, for each bound region to functionally test in transcriptional assay in my study, I cloned a full conserved genomic element into the luciferase-containing vector. As a result, the cloned region of DNA is longer than the size of the bound region identified by ChIP-seq for each putative enhancer. This reduced the likelihood of disrupting the enhancer by missing adjacent sequences that make up the entire regulatory element. The length of the genomic regions cloned in the luciferase reporter vector is shown in Table 3.2. Ben Martynoga assigned a PhastCons score to the conserved sequences with the PhastCons program (Siepel et al., 2005). PhastCons program identifies evolutionary conserved elements in a multiple alignment, given a phylogenetic tree. It assesses similarity/divergence of aligned sequences among species using statistical models of nucleotide substitution and generates a score that ranges between 0 and 1, with 0 for non-conserved regions and 1 for conserved regions.

PhastCons scores for each cloned region are shown in Table 3.2. As shown in the table, the analysis of the conservation demonstrated that among all regions, only 3 out of 9 (MSB25, MSB24, and MSB26) exhibit the highest level of conservation. Region MSB 18 shows modest level of conservation with a PhastCons score of 0.271. The remaining five regions (MSB4, MSB10, MSB11, MSB22, and MSB23) are only minimally conserved (Table 3.2). PhastCons diagrams from the UCSC genome browser in Figure 3.5 show the area of conservation of genomic sequences among vertebrate species for the least conserved and the most conserved regions identified (Figure 3.5).

In conclusion, the putative regulatory elements identified in this study showed little sign of evolutionary constraint, with only 3 regions out of 9 being highly conserved.

Genomic Region - (ID name)	Cloned Region - Size in bp	PhastCons Score - (Sequence conservation)
MSB 4	336 bp	0.039
MSB 10	616 bp	0.169
MSB 11	495 bp	0.057
MSB 18	874 bp	0.271
MSB 22	723 bp	0.147
MSB 23	611 bp	0.154
MSB 24	678 bp	0.475
MSB 25	500 bp	0.555
MSB 26	774 bp	0.456

Table 3.2. Sequence conservation and size of the putative regulatory elements cloned in luciferase vector.

For each element bound by Ascl1 and Sox2 and selected as putative regulatory element, a full conserved genomic region was cloned in the luciferase reporter vector to assess its transcriptional function. The size of the genomic regions cloned are shown in the table. Evolutionary conservation of the regions in the genome of vertebrate species has been analysed using PhastCons program. A PhastCons score was assigned to each region, as shown in the table. The regions selected exhibited little sign of evolutionary constraint, with only 3 regions out of 9 being conserved: MSB 25, the most conserved, MSB 24, and MSB 26. MSB 18 exhibited a moderate conservation. All other 5 regions were only minimally conserved.

MSB 4 - 336bp - PhastCons score 0.039**MSB 25 - 500bp - PhastCons score 0.555****Figure 3.5. PhastCons diagram of sequence conservation.**

The sequence conservation among vertebrate species of the putative regulatory regions are shown in PhastCons diagrams for the least conserved region (MSB4) and the most conserved region (MSB25). For each region, the PhastCons score of conservation and the size of the genomic element cloned in the β -globin luciferase vector to test in luciferase assay is shown.

3.3.4 Recruitment of p300 on the bound regions: a hallmark for the identification of genomic enhancers

Conservation of non-coding genomic sequences has proved to be limited in the identification of regulatory elements, as reported in literature (Visel et al., 2009b, Blow et al., 2010). As alternative, specific chromatin signatures have proved to be a valuable hallmark in the identification of genomic functional enhancers and other non-coding regulatory elements (Heintzman et al., 2007, Heintzman et al., 2009, Mikkelsen et al., 2007, Creighton et al., 2010). The recruitment of the transcriptional coactivator p300 is among these chromatin marks (Visel et al., 2009a, Blow et al., 2010, Visel et al., 2009b, Ghisletti et al., 2010).

The histone acetyltransferase and transcriptional coactivator p300 is a component of enhancer-associated protein assemblies. It has been shown to accurately predict the genomic location and activity *in vivo* of developmental enhancers (Visel et al., 2009b, Visel et al., 2009a, Blow et al., 2010). The protein p300 is thought to increase gene expression mainly in three ways: by relaxing the chromatin structure at the gene

promoter through its intrinsic histone acetyltransferase activity; recruiting the basal transcriptional machinery including RNA polymerase II to the promoter; and acting as transcriptional adaptor molecule, which mediates specific protein-protein interactions and bridges the proteins on the enhancers with the transcriptional machinery on the promoter (Vo and Goodman, 2001).

Therefore, to further confirm the nature of regulatory sequences for the candidate bound regions, we performed a ChIP experiment, using antibody against the protein p300. Primers flanking the bound regions (Table 2.1 in Materials and Methods) were used for amplification of these elements in chromatin from NS5 cells and enrichment of the precipitated regions was quantified by ChIP-qPCR. Dll1 ORF served as a negative control (on grey background in Figure 3.6), whereas Hom2 and Nes258 served as positive controls (on violet background in Figure 3.6), as known enhancers activated by Ascl1 and Sox2 (Tanaka et al., 2004). The experiment showed enrichment for p300 on all nine genomic regions tested, with fold enrichment values between 2 and 8 times that of the negative control. However, also the mock chromatin exhibited high values for each region. Nevertheless, the enrichment values for p300 are between 2 and almost 6 times the mock values for all 9 regulatory regions analysed, supporting the idea that true binding of p300 occurs on these regions (Figure 3.5).

In conclusion, we could confirm recruitment of p300 on all nine selected genomic elements, further suggesting their regulatory functions. Besides recruitment of p300, ChIP-seq data from Ben Martynoga also showed binding of the histone mark H3K27Ac on the bound regions supporting their nature of active enhancer (data not shown) (Martynoga et al., 2013).

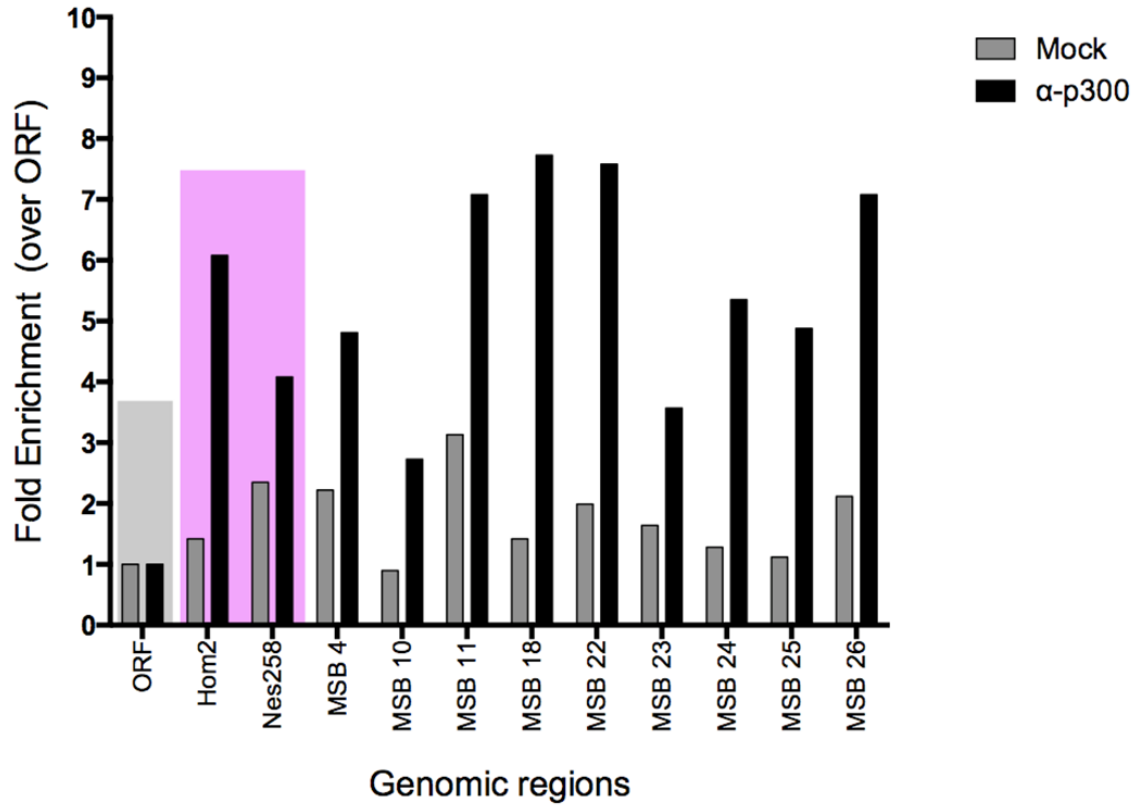


Figure 3.6. Recruitment of the transcriptional coactivator p300 on the putative regulatory regions.

ChIP-qPCR has been performed to quantify enrichment of the protein p300 on the putative regulatory regions in chromatin from NS5 cells immunoprecipitated with α -p300 antibody compared to mock immunoprecipitated chromatin from the same NS5 cells. All 9 regions tested were enriched and bound by p300 (fold enrichment over the ORF, negative control). Negative control and positive controls on grey and violet background respectively.

3.4 Transcriptional activity of the genomic regions bound by *Ascl1* and *Sox2*

To test the regulatory functions of the nine genomic sequences, we performed transcriptional reporter assays. Each genomic element was cloned upstream of the β -globin minimal promoter driving basal expression of the firefly *luciferase* gene, as described in paragraph 2.3.1 in Materials and Methods.

NS5 cells were co-transfected with each enhancer-containing luciferase vector and a constitutive *lacZ* reporter construct to normalise for transfection efficiency. Cells were harvested 24 hours after transfection to measure luciferase activity. Luciferase and β -galactosidase assays were performed in triplicates or quadruplicates and results presented as mean \pm standard deviations (SD). The results shown are representative of two or three independent repeats.

3.4.1 Basal activity of the bound regions: regulation by endogenous TFs in NS5 cells

Figure 3.7 shows that all regions, except for MSB4 and MSB10, increased basal luciferase expression compared to the activity of the enhancer-less vector (empty vector). The increase ranged between 4-fold (for MSB22) and 16-fold (for MSB18) the activity of the empty vector. Therefore, the selected genomic sequences showed regulatory functions, *in vitro*, with characteristics of enhancers in neural stem cells. It is possible that they are activated by endogenous TFs in neural stem cells. Only regions MSB4 and MSB10 did not show basal enhancer activity induced by endogenous TFs.

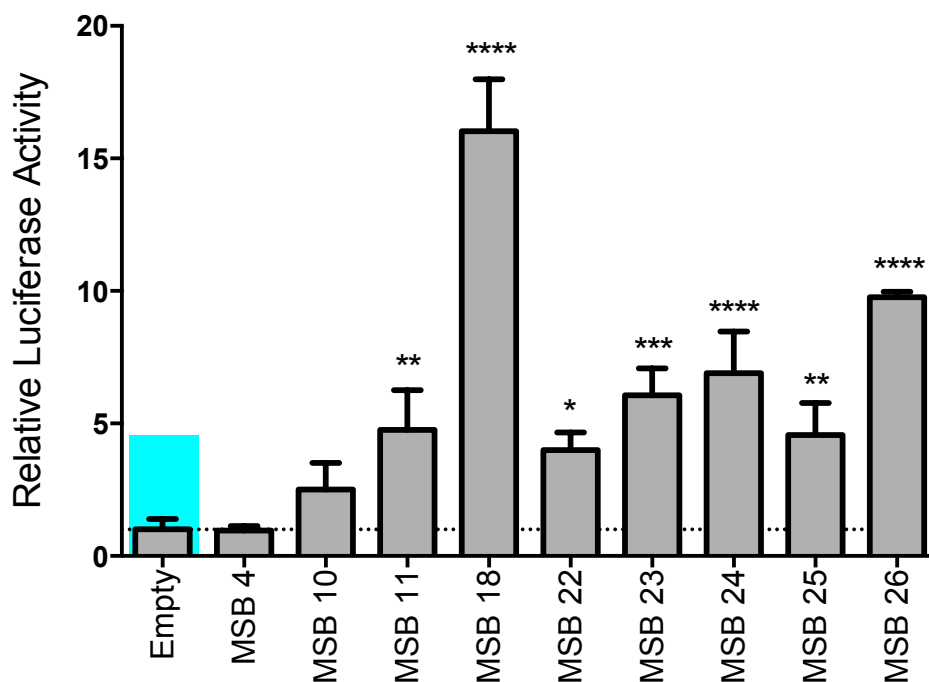


Figure 3.7. Transcriptional activity of the bound regions in neural stem cells by luciferase assay.

Luciferase activity of the regions is expressed as fold change over the activity of the enhancer-less, empty vector used as negative control, on light blue background. Almost all regions show basal enhancer activity in NS5 cells. Luciferase activity values for almost all regions are between 4 and 16 fold the activity of the empty vector, with value of 1. Regions MSB4 and MSB10 do not show luciferase activity. Data are presented as the mean \pm SD of quadruplicate assays. The experiment shown is representative of two independent biological replicates.

3.4.2 Regulation of enhancer activity by Ascl1 and Sox2

To determine the effect of TFs Ascl1 and Sox2 on the regulation of these elements, we tested enhancer activity by cotransfection experiments in NS5 cells. The expression vectors pCAGGs-Ascl1 and pCAG-Sox2 were cotransfected into NS5 cells with the enhancer-containing luciferase vectors, either alone or together. Cells were harvested 24 hours after transfection to measure luciferase activity. In these experiments, the activity of each enhancer induced by Ascl1 and/or Sox2 overexpression was compared to the activity of the same enhancer induced by GFP expression using pCAGGs-GFP as a control.

As shown in Figure 3.8 A and C, exogenous *Ascl1* increased the activity of seven enhancers compared to the GFP expression vector condition (MSB4, MSB18, MSB22, MSB26 in Figure 3.8A; MSB10, MSB11, and MSB25 in Figure 3.8C) (red bars in the graphs). Similarly, exogenous *Sox2* increased the activity of five enhancers (MSB23, MSB24 in Figure 3.8B; MSB10, MSB11, and MSB25 in Figure 3.8C) (green bars in the graphs), but in this case the activation was moderate. MSB24 and MSB25 revealed the highest activation by *Sox2* (Figure 3.8B and C, respectively). Finally, when *Ascl1* and *Sox2* were overexpressed simultaneously in NS5 cells, six regions showed activation (MSB4, MSB18 in Figure 3.8A; MSB24 in Figure 3.8B; MSB10, MSB11, and MSB25 in Figure 3.8C) (purple bars in the graphs).

Interestingly, for almost all enhancers tested, the up-regulation by the two TFs overexpressed together was lower than that shown by the overexpression of each single TF, either *Ascl1* or *Sox2*. This result suggests that the two TFs, *Ascl1* and *Sox2*, counteract each other in the regulation of most co-bound regulatory regions (MSB4, MSB18, MSB22, MSB26 in Figure 3.8A; MSB23, and MSB24 in Figure 3.8B). More precisely, luciferase reporter assays revealed a potent activation of the regulatory regions MSB4, MSB18, MSB22 and MSB26 by *Ascl1* (Figure 3.8A). However, when *Sox2* was overexpressed together with *Ascl1*, *Sox2* was able to block most of *Ascl1*'s ability to induce luciferase activity, suggesting that *Sox2* antagonizes *Ascl1* on these elements. It is possible that *Sox2* counteracts *Ascl1* in the activation of the target genes by competing with the latter TF for shared co-activators or general transcription factors (GTFs) when the two TFs are overexpressed together in the NS5 cells, and therefore as an effect of transcriptional squelching. Transcriptional squelching occurs when increasing concentrations of transcriptional activators paradoxically inhibit the transcription of a target gene by titrating one or more GTFs or co-activators in limited

supply within the host cell. This mechanism occurs particularly on episomal DNA and it is largely observed in transient transfection assays (Natesan et al., 1997). Overexpression of Sox2 alone did not have any effect on the activity of these four enhancers (MSB4, MSB18, MSB22 and MSB26). Conversely, two regions, MSB23 and MSB24, were upregulated by Sox2 and either not induced (MSB23) or repressed (MSB24) by Ascl1 in single overexpression of each TF (Figure 3.8B). In this case, when Ascl1 was overexpressed together with Sox2, Ascl1 was able to block most of Sox2's ability to induce luciferase activity. Thus, Ascl1 antagonised Sox2 on these enhancers. Also in this case the mechanism of counteraction between the two TFs can result from transcriptional squelching associated to the experimental conditions of transient transfection assays. Finally, on three other regions, MSB10, MSB11, and MSB25, the two TFs had a neutral interaction, where the activity in the single and double overexpression was very similar (Figure 3.8C). Evidence of synergistic activity was not seen on any of the tested enhancers (Figure 3.8A, B, and C). Taken together, these findings suggest that Ascl1 and Sox2 act as independent regulators of the enhancers identified in this study, and antagonise each other in the regulation of 6 of the 9 enhancers characterised.

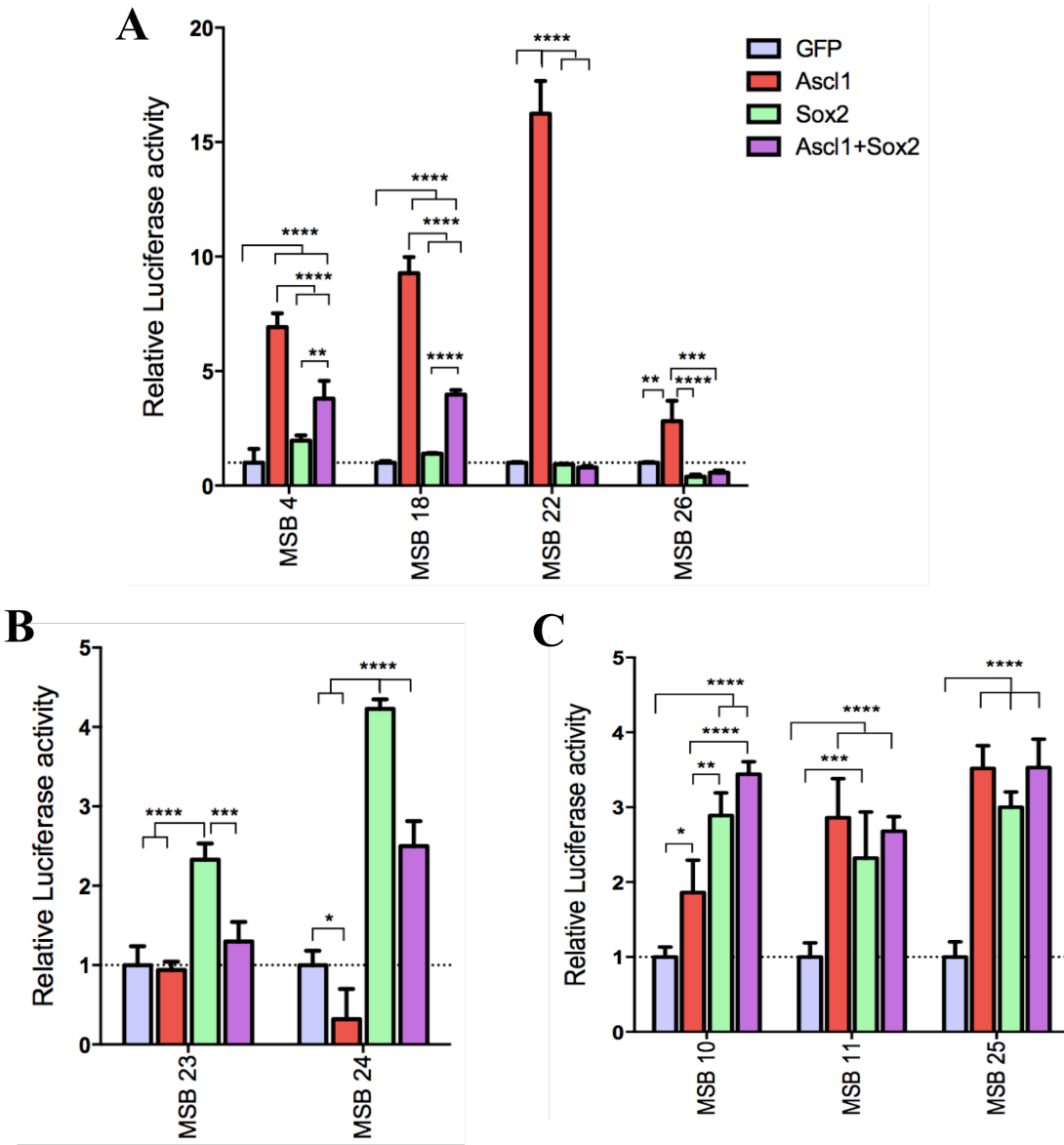


Figure 3.8. Interaction of Ascl1 and Sox2 in the regulation of the neural enhancers.

Neural enhancers were transfected in NS5 cells together with expression vectors for Ascl1 and Sox2 overexpressed either alone or together. Luciferase activity was measured after 24 hours. For each enhancer, luciferase activity is expressed as relative to the activity of the GFP expression vector used as control. All 9 enhancers were regulated by exogenous Ascl1 and Sox2. Enhancers in group A were strongly activated by exogenous Ascl1 while Sox2 counteracted Ascl1-induced activation when the 2 TFs were simultaneously overexpressed in NS5 cells (A). Enhancers in group B were activated by exogenous Sox2 whereas Ascl1 counteracted Sox2-induced activation when the 2 TFs were simultaneously overexpressed (B). Enhancers in group C showed a neutral interaction between the 2 TFs with the activity by exogenous Ascl1 and/or Sox2 being very similar when the 2 TFs were overexpressed either alone or together (C). Ascl1 and Sox2 counteract each other in the regulation of 6 out of 9 enhancers (A and B). Synergistic activation by the 2 TFs was not observed on any of the enhancers identified. Data are presented as the mean \pm SD of triplicate assays. The experiment shown is representative of three independent biological replicates.

3.4.3 Classification of the enhancers: neural stem cells and neuronal enhancers

Cotransfection experiments demonstrated that Ascl1 and Sox2 counteract each other in the regulation of almost all enhancers identified. These experiments also revealed that two different patterns of counteraction occur for different enhancers, with some activated by Ascl1 and inhibited by Sox2, and others activated by Sox2 and repressed by Ascl1. Given that distinct patterns of interaction for the nine enhancers identified are possible, we attempted a classification of the enhancers to clarify and understand the nature of the transcriptional interaction between Ascl1 and Sox2. The effect of the overexpression of the TF Ascl1 on the cellular and molecular context in the neural stem cells was also addressed in the classification. It is known that overexpression of Ascl1 in neural stem cells induces neuronal differentiation (Lee et al., 1995, Farah et al., 2000). After 24 hours of Ascl1 overexpression, at the time when luciferase activity was measured in my experiments, NS5 cells have acquired characteristics of neurons (Martynoga, unpublished data, and data not shown). Therefore, we classified the enhancers activated by overexpression of Ascl1 as neuronal enhancers active in differentiating neurons. According to this, we distinguished enhancers active in neural

stem cells from neuronal enhancers. The first category includes enhancers that exhibit basal luciferase activity in NS5 cells, therefore neural stem cell active enhancers (Figure 3.9A). The second category includes enhancers activated by overexpression of *Ascl1* in NS5 cells, therefore neuronal enhancers (Figure 3.9C). Figure 3.9 illustrates this classification made on the basis of basal activity and/or activation by exogenous *Ascl1* shown by the enhancers in luciferase assay experiments. Enhancers MSB23 and MSB24 are examples of enhancers active in neural stem cells (Figure 3.9 A). They show basal activity in NS5 cells, but are not activated by exogenous *Ascl1*. Enhancers MSB4 and MSB10 are examples of neuronal enhancers (Figure 3.9 C). They are activated by exogenous *Ascl1* but do not show basal activity in NS5 cells. The remaining 5 enhancers MSB11, MSB18, MSB22, MSB25, and MSB26 show both basal activity in NS5 cells and activation by exogenous *Ascl1*. Therefore they have been classified as both neural stem cell active enhancers and neuronal enhancers (Figure 3.9 B).

The effect of the overexpression of *Sox2* on the regulation of the enhancers was also considered and added to the classification. In this way, we identified and distinguished enhancers activated by exogenous *Sox2* from enhancers inhibited by exogenous *Sox2* (where *Sox2* inhibited *Ascl1* activity when the two TFs were overexpressed together). This second classification is shown in Figure 3.10. The first group includes MSB23, MSB24, MSB10, MSB11, and MSB25. These enhancers are either neural stem cell active enhancers or NSC/Neuronal enhancers (Figure 3.10 A and B). The second group includes the remaining enhancers shown in Figure 3.10 C. These enhancers are NSC active/Neuronal enhancers. This group also includes the neuronal enhancer MSB4. Table 3.3 summarizes the entire classification of the enhancers.

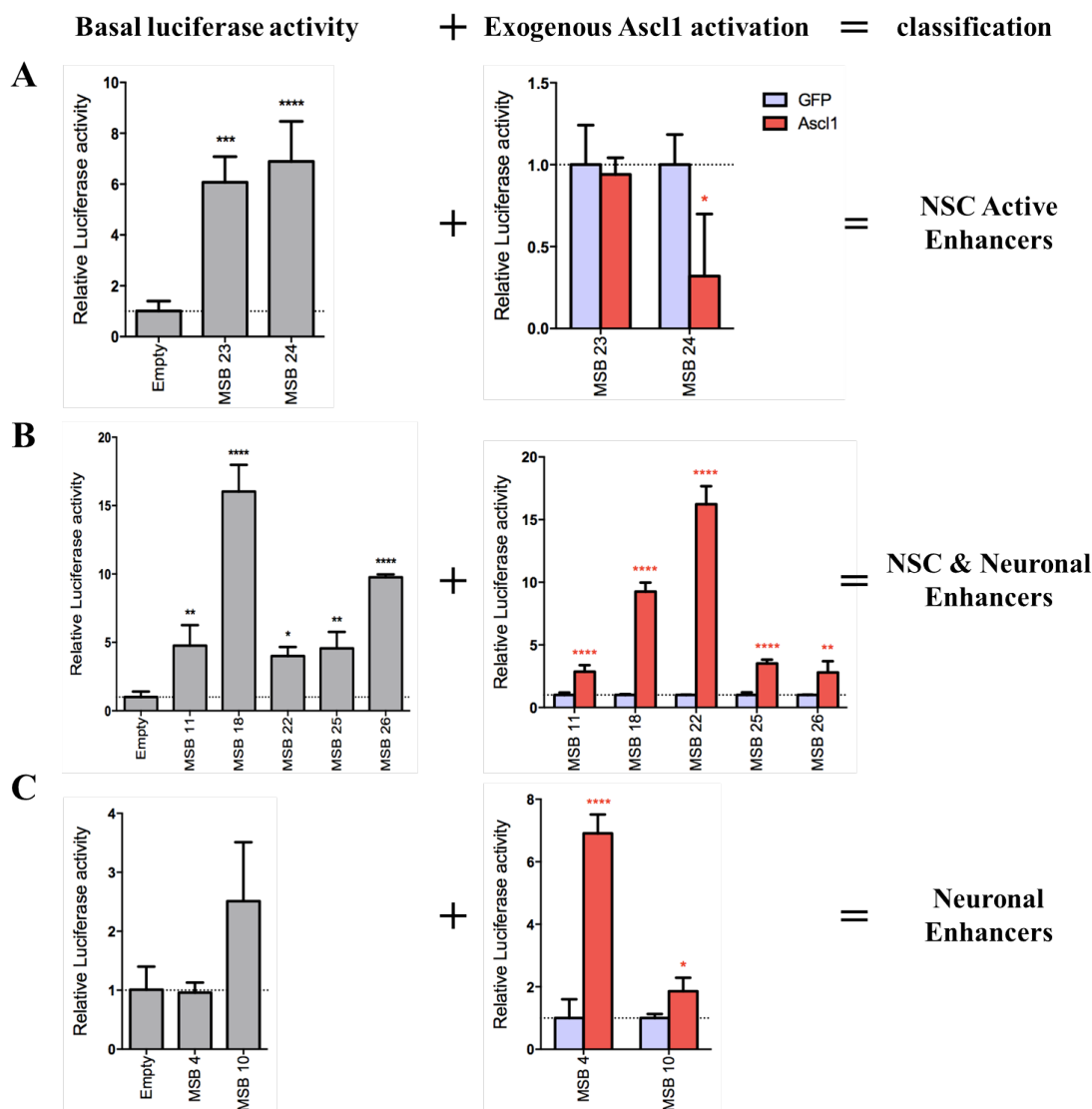


Figure 3.9. Classification of neural stem cells and neuronal enhancers.

Enhancers have been classified on the basis of the exogenous activation of Ascl1 and the basal enhancer activity in neural stem cells. (A) Enhancers with basal luciferase activity in NS5 cells have been classified as enhancers active in neural stem cells. (C) Enhancers activated by exogenous Ascl1 have been classified as neuronal enhancers, due to the neuronal differentiation induced by overexpression of Ascl1 in NS5 cells. (B) Enhancers displaying basal luciferase activity in NS5 cells and activated by exogenous Ascl1 have been classified as both enhancers active in NSCs and neuronal enhancers. (NSC=Neural Stem Cell)

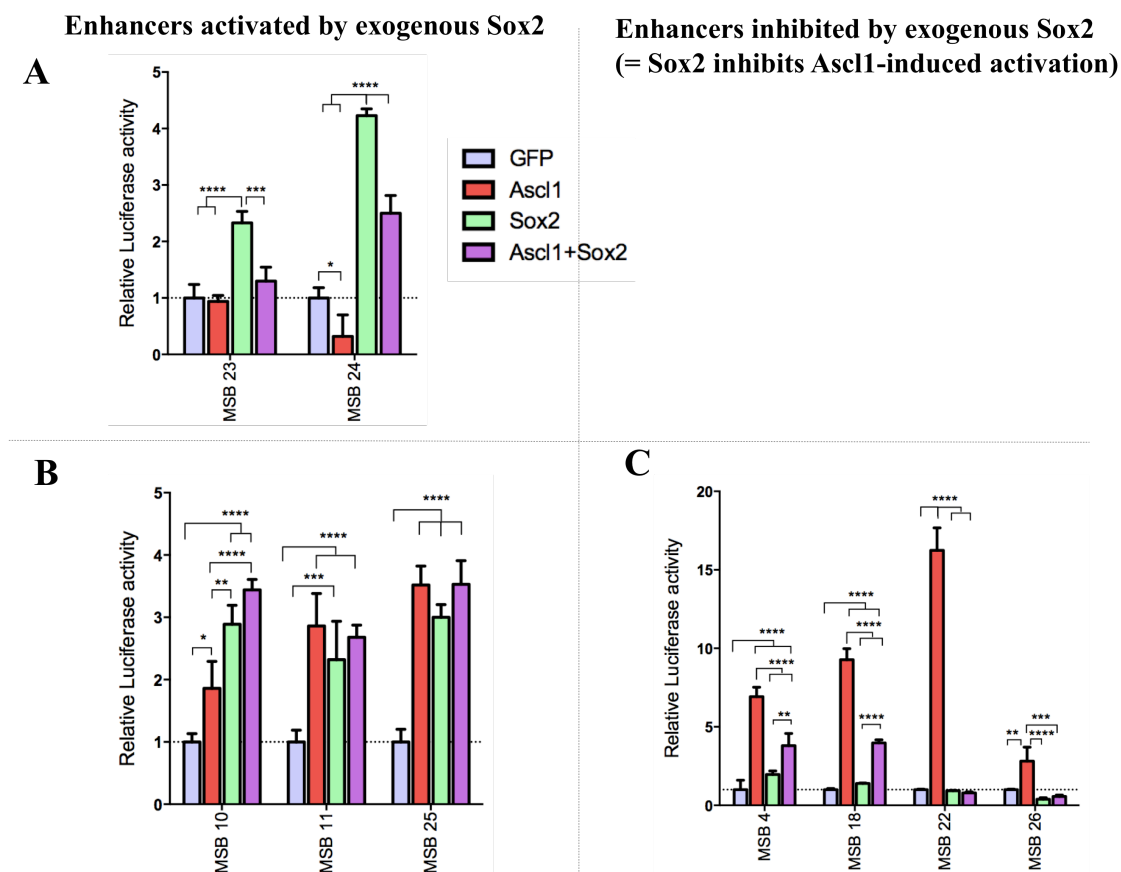


Figure 3.10. Classification of the enhancers activated by exogenous Sox2 or inhibited by exogenous Sox2.

Enhancers in both group A and B are activated by exogenous Sox2. Enhancers in group A are NSC active enhancers only, according to previous classification in Figure 3.9A. Enhancers in group B are both NSC active and neuronal enhancers. For the enhancers in group C, Sox2 inhibits Ascl1-induced activation.

	Enhancers activated by Sox2	Enhancers inhibited by Sox2 - (=Sox2 inhibits Ascl1-induced activation)
NSC only (basal luciferase activity)	<ul style="list-style-type: none"> • MSB23 • MSB24 	
NSC &/or neuronal (basal activity & activation by exog. Ascl1)	<ul style="list-style-type: none"> • MSB10 • MSB11 • MSB25 	<ul style="list-style-type: none"> • MSB4 • MSB18 • MSB22 • MSB26

Table 3.3. Classification of enhancers.

The Table summarizes the full classification of the enhancers. These have been divided as NSC active only or NSC &/or neuronal (rows) if they only display basal luciferase activity in NS5 cells (first group) or also activation by exogenous Ascl1 in NS5 cells. This first classification has been done to consider the effect of exogenous Ascl1 in NS5 cells, which induces neuronal differentiation and therefore changes the cellular and molecular context of the NS5 cells. The enhancers have also been divided as activated by exogenous Sox2 or inhibited by Sox2 (= Sox2 inhibits Ascl1-induced activation) (columns).

3.5 Regulation of the neural enhancers by Ascl1-EnR and Sox2-EnR

To further confirm the role of Ascl1 and Sox2 in the regulation of the genomic elements through direct DNA binding, dominant repressor forms of the 2 TFs were also transfected in NS5 together with enhancer-containing vectors and tested in the luciferase assay. The dominant repressor forms, Ascl1-EnR and Sox2-EnR, as mentioned earlier, consist of the DNA binding domains of the respective TF fused to the engrailed repressor domain, which represses the transcriptional activity of the bound region. Also in this case the activity of each enhancer in the presence of the dominant repressor forms, Ascl1-EnR or Sox2-EnR, was compared to the activity of the same enhancer driven by GFP expression vector used as control. As expected, all nine enhancers were repressed by Sox2-EnR (Figure 3.11). Almost all enhancers were also repressed by Ascl1-EnR. However, enhancers MSB11 and MSB24 showed no repression by Ascl1-EnR, suggesting that Ascl1 contributes little to their basal activity

(Figure 3.11). In conclusion, these experiments suggest that the regulation of enhancer activity is driven through direct binding of Ascl1 and Sox2, confirming the role of the two TFs in the regulation of the enhancers here characterised.

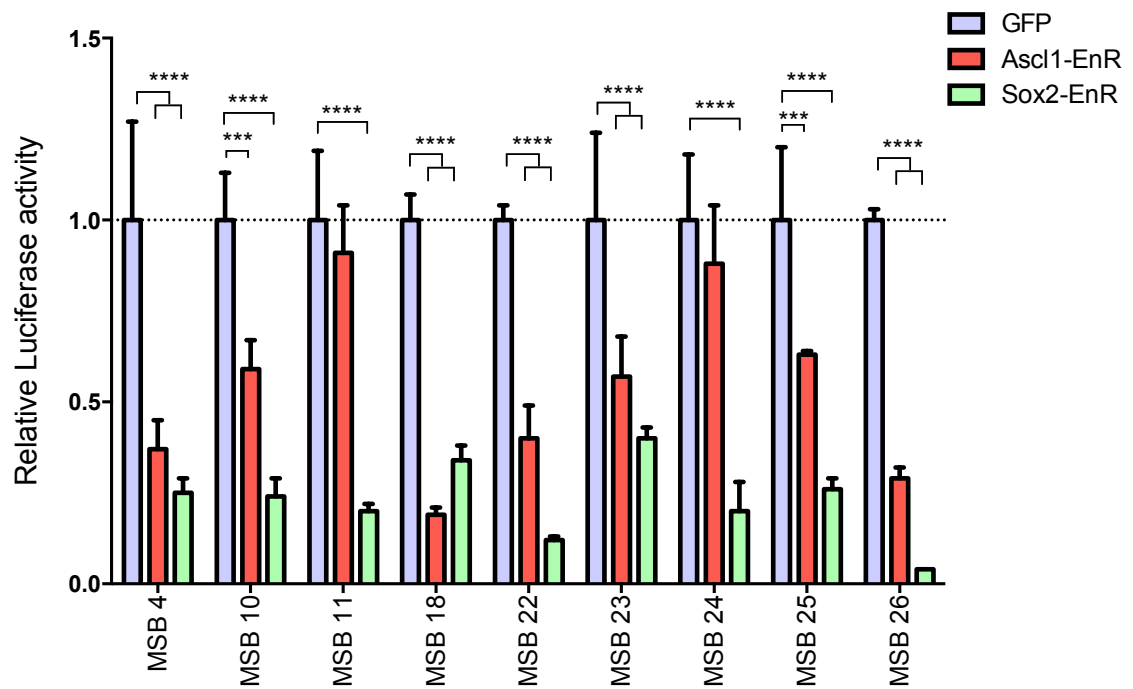


Figure 3.11. Repression of the enhancer activity by Ascl1-EnR and Sox2-EnR.

Each enhancer was transfected in NS5 cells together with Ascl1-EnR or Sox2-EnR expression vectors. Transcriptional activity of the enhancers was measured. All enhancers were repressed by Sox2-EnR compared with the activity of the GFP negative control. Enhancers were repressed by Ascl1-EnR with the exception of enhancers MSB11 and MSB24. Results of this experiment suggest that Ascl1 and Sox2 regulate the neural enhancers through their DNA binding domains and direct binding to the regulatory elements.

3.6 Discussion

The objective of part of this study was the identification and *in vitro* characterisation of neural enhancers regulated by the TFs Ascl1 and Sox2. Moreover, this study also aimed to dissect the nature of the possible interaction between Ascl1 and Sox2 as transcriptional regulators of co-bound genomic elements.

We started this study with the identification by ChIP-seq of genomic regions where Ascl1 and Sox2 could bind in close proximity, suggesting their recruitment on common genomic elements. This molecular data set us to investigate any possible interaction between the two proteins.

The computational analysis and the results described in this chapter had primarily the purpose to validate the ChIP-seq data and select genomic elements with characteristics of regulatory regions. The regions selected were co-bound by Ascl1 and Sox2, according to ChIP-seq data. They were associated to genes de-regulated by perturbation of Ascl1 and Sox2 according to microarray data. The genes associated to the regions were also expressed in the telencephalon of the developing CNS, and some of them in the same expression domain of Ascl1 and Sox2 (Figure 3.2). These features strengthen the possibility of a functional regulation by Ascl1 and Sox2.

The sequence conservation of the putative regulatory regions among vertebrate species was also considered, since comparative genomic approaches have led to successful identification of many functional enhancers. These approaches are based on the notion that the sequences of regulatory elements tend to be evolutionary constrained among different species, due to the deleterious consequences of their changes (Visel et al., 2009b). However, in our case the PhastCons analysis revealed that the majority of the identified genomic sequences were poorly conserved, with only 3 out of 9 elements having score close to 0.5, a medium value of conservation. Indeed, it has been emerging that genomic constraint of regions could suffer from some limitations in the identification of enhancers. A large part of them are not evolutionary constrained. Interestingly, a work by Blow et al. characterized functional active heart enhancers weakly conserved among vertebrate species (Blow et al., 2010), identifying a large population of developmental enhancers escaping negative evolutionary selection.

Recruitment of the transcriptional coactivator p300 has been shown to accurately predict the genomic location of active enhancers (Visel et al., 2009a, Blow et al., 2010, Ghisletti et al., 2010). Therefore, we analysed the recruitment of this transcriptional coactivator on the selected bound regions to strengthen the assumption that they function as regulatory regions. ChIP experiment for the protein p300 and ChIP-qPCR showed binding on all nine selected genomic elements, suggesting their nature of enhancers. This result is also consistent with a p300 ChIP-seq dataset obtained in the lab, where peaks were called on the same genomic elements tested here (Martynoga et al., 2013).

At a molecular level, by ChIP-qPCR we could prove that 9 out of the 14 selected bound regions were enriched in chromatin from NS5 cells immunoprecipitated with α -Ascl1 and α -Sox2 antibody, suggesting binding of the two TFs on the same genomic elements. At a functional level, we were able to prove that these elements acted as enhancers regulated by Ascl1 and Sox2. The elements were tested by luciferase assays and most exhibited basal enhancer activity *in vitro* in NS5 cells, with the exception of MSB4 and MSB10, possibly responding to the regulation by endogenous TFs in neural stem cells. When testing the effect of the overexpression of Ascl1 and/or Sox2 on the activity of the enhancers, they were all responding to their regulation, although in different ways. More precisely, 7 out of 9 regions were activated by Ascl1 and 5 out of 9 were activated by Sox2. However Ascl1 acted as a strong positive regulator of the enhancers, increasing the activity of the regions between 2 and 16 fold the activity of the GFP control vector. Only MSB24 was repressed by Ascl1. The activation by Sox2 was often very modest instead. MSB24 and MSB25 showed the highest activation by Sox2, but this never exceeded 4 times the activity of GFP control vector. Therefore, it can be concluded that most enhancers were more responsive to exogenous Ascl1 than

Sox2, besides the suggested recruitment and binding of both TFs on the regions characterised.

The other interesting result emerging from the transcriptional assay was that Ascl1 and Sox2 counteracted each other in the regulation of 6 out of 9 enhancers. Specifically, Sox2 inhibited Ascl1 activation on MSB4, MSB18, MSB22, and MSB26 which were also the enhancers showing the highest activation by Ascl1. Conversely, Ascl1 blocked Sox2 activation on enhancers MSB23 and MSB24, with this last one being the enhancer activated most potently by Sox2. Mechanisms of transcriptional squelching might account for the counteraction observed when the two TFs are overexpressed together in co-transfection experiments. A synergistic interaction between the two TFs was never seen on the enhancers identified in this study.

Since the effect of Ascl1 and Sox2 on the regulation of the enhancers was tested by overexpression in NS5 cells, the effect of their gain of function on the neural stem cell fate needed to be considered. In this study, the effect of the simultaneous overexpression of Ascl1 and Sox2 on the neural stem cell fate *in vivo* was not evaluated. However, it is known that gain of function of bHLH proneural proteins, including Ascl1, can induce a neuronal fate in undifferentiated progenitor cells (Lee et al., 1995, Farah et al., 2000). NS5 cells start to differentiate into neurons already 24 hours after transfection of Ascl1 (Martynoga, unpublished data). Therefore, we classified the enhancers as active in neural stem cells or neuronal enhancers if they displayed basal activity in NS5 cells or were induced by exogenous Ascl1, respectively. Some enhancers were active in both cellular and molecular context, therefore were classified as both active in neural stem cells and as neuronal enhancers when induced by overexpression of Ascl1.

In conclusion, the results described in this chapter have identified a group of functional neural enhancers *in vitro* among genomic regions identified through a ChIP-seq approach. The TFs Ascl1 and Sox2 were able to regulate the identified enhancers, as predicted by their recruitment on the sequences. However, Ascl1 was the strongest activator of these elements, whereas Sox2 showed only a modest activation on a smaller number of the identified elements. Intriguingly, Ascl1 and Sox2 counteract each other in the regulation of most of these enhancers and never appear to interact synergistically in the regulation of the regions where they are commonly recruited. Finally, as shown from the classification of the 9 enhancers characterised in this study, even within a small group of enhancers there is much diversity of regulation, suggesting rather complex interactions between Ascl1 and Sox2.

The identification of TFBSs (Transcription Factor Binding Sites) for Ascl1 and Sox2 and their involvement in driving the enhancer activity was necessary for further characterisation of these enhancers and their mechanisms of transcriptional regulation. Such work is described in the next chapter.

4. Identification of the E-boxes and Sox binding motifs and different mechanisms of regulation of the neural enhancers

4.1 Introduction

In the previous chapter I demonstrated that the TFs Ascl1 and Sox2 act as regulators of the neural enhancers identified. In this chapter, I identified Transcription Factor Binding Sites (TFBSs) or Binding Motifs (BM)s for Ascl1 and Sox2 in the sequence of the enhancers to assess their role in the enhancer activity and to infer if the regulation occurs through direct DNA binding of the TFs to their motifs. To evaluate the function of the consensus motifs, I introduced point mutation in their sequences to assess if the activity of the enhancers was compromised when TF binding motifs were lost. Transcriptional assays of mutant and wt enhancers were performed to answer to this question.

4.2 E-boxes and Sox binding motifs in the neural enhancers

Transcription factors binding motifs were identified using the TESS software (Transcription Element Search System) (<http://www.cbil.upenn.edu/cgi-bin/tess/tess>). A list of E-boxes and Sox consensus motifs searched within the enhancer sequences is shown in Table 4.1. More precisely, two known E-boxes, E1-box and E2-box, were searched as Ascl1 binding motifs (Castro et al., 2006). Several known motifs were searched for Sox factors, which all exhibit a preference for the hexameric core sequence 5'-WWCAAW-3', where W indicates A or T (Lefebvre et al., 2007). In particular, I searched for the following Sox consensus motifs using the TESS software:

the Sox motif identified in the *Nestin* enhancer and bound by Sox2, as reported by Tanaka and colleagues, thus named as Sox Nestin in Table 4.1 (Tanaka et al., 2004); a Sox 5 motif obtained from the Jaspar database (Sandelin et al., 2004); two Sox motifs, named Sox b and Sox trawl-a in Table 4.1, obtained by Ben Martynoga in a *de novo* motif analysis of Sox2 ChIP-seq data and using the Trawler database (Ettwiller et al., 2007). All 9 enhancers characterized in this project have at least one binding motif for Ascl1 and one binding motif for Sox2. However, no clear similarities in the architecture of the enhancers were observed, in term of motif numbers, order, positioning, and spacing. The binding motifs identified in the enhancer sequences are shown in Figure 4.1. E-boxes are shown as red triangle and Sox motifs in green. The purple rectangle shows the region where Ascl1 and Sox2 ChIP-seq peaks overlap within the enhancer, according to ChIP-seq data, therefore the region where the two TFs bind together in close proximity.

BM's consensus list	Consensus sequence	Mutated sequence	References
E1-box	<u>CAGG</u> TG	<u>TAGG</u> <u>CT</u>	Castro et al., 2006
E2-box	<u>CAGC</u> TG	<u>TAGC</u> <u>CT</u>	Castro et al., 2006
Sox Nestin	G <u>ACAAAA</u>	G <u>GTGCTC</u>	Tanaka et al., 2004
Sox 5	A <u>ACA</u> AT	A <u>GTGCT</u>	Jaspar database
Sox b	C <u>ACA</u> ATG	C <u>GTGCT</u> G	Martynoga, unpubl.
Sox trawl-a	<u>ACAA</u> AG	<u>GTGCT</u> G	Trawler database

Table 4.1. Binding Motifs consensus sequences within neural enhancers.

The BM's consensus sequences listed in the table were searched within the neural enhancers using the TESS software to identify E-box and Sox motifs for Ascl1 and Sox factors. The mutated sequences for each consensus are also shown in the table. The positions and the nucleotide substitutions are shown in red and are underlined. Consensus motifs were obtained from the literature and/or from on line databases, such as Jaspar and Trawler, as indicated by the references. The two E-boxes described in *Castro et al., 2006* were mutated also as according to *Castro et al., 2006*. Sox motifs were all mutated in a similar way, according to the nucleotide substitutions of the *Nestin* core sequence introduced in *Tanaka et al., 2004*.

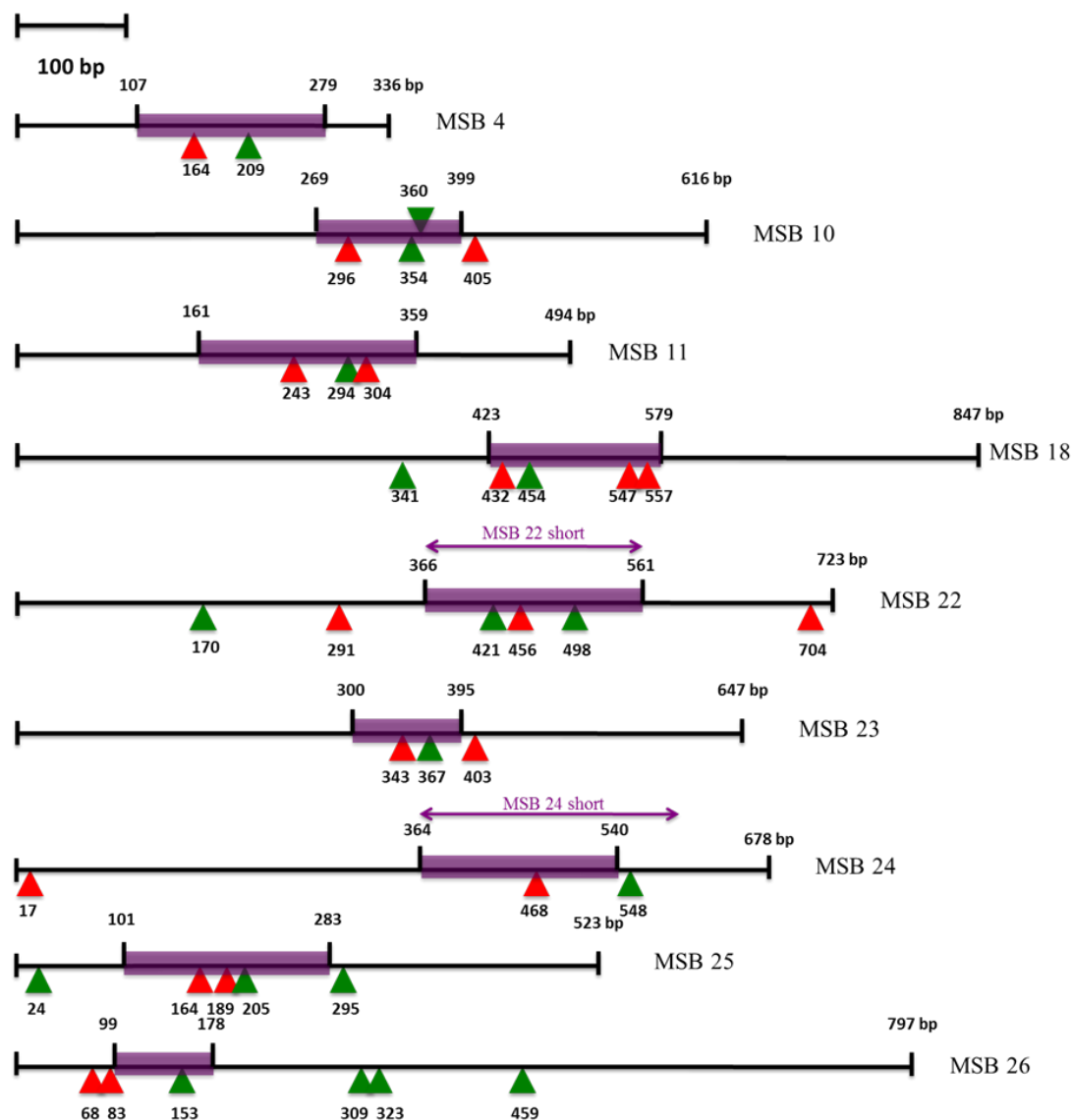


Figure 4.1. E-boxes and Sox motifs within the neural enhancers.

Segments represent the enhancer sequences. The sequence length is shown for each enhancer. E-boxes and Sox motifs are represented as red and green triangles respectively, as shown in the legend. Regions where Ascl1 and Sox2 ChIP-seq peaks overlap are shown as purple rectangles, as also shown in the legend. The numbers indicate the nucleotide positions in the cloned sequence of the enhancer, and they indicate the position of the first nucleotide of the binding motif for the E-boxes and the Sox motifs, under the red and green triangle respectively. For instance, enhancer MSB4 has a sequence of 336 base pair. 164 and 209 indicate the position of the first nucleotide of the E-box and Sox motif in the sequence, respectively. Therefore the distance between the two binding motifs is 39 bp (209-164-6bp). The region where Ascl1 and Sox2 ChIP-seq peaks overlap is between nucleotide 107 and 209 in the cloned sequence of the enhancer.

To assess if TFs regulate enhancer activity through direct binding to the DNA consensus motifs, I introduced point mutation in the motif sequences known to disrupt the binding (Tanaka et al., 2004, Castro et al., 2006). In particular, mutations disrupted the E-boxes so that Ascl1 should not bind the sequence. Nucleotide substitutions were introduced in the two E-boxes as described in Castro et al. (Castro et al., 2006). All Sox motifs identified were mutated in the same way so that any Sox factor should not bind the sequence. Nucleotide substitutions were introduced in the core positions of the Sox motifs as described for the *Nestin* enhancer in Tanaka et al. (Tanaka et al., 2004). Table 4.1 shows all point mutations introduced in the binding motifs. The transcriptional activity of the mutant and wt enhancers were tested and compared in transcriptional assays *in vitro* in NS5 cells, with the same experimental approach previously described for the wt enhancers.

The following enhancers were mutated, choosing at least one from each class of neural stem cell active or neuronal enhancers, with different effects of exogenous Ascl1 and exogenous Sox2 activity: enhancers MSB4 (neuronal enhancer), MSB18, and MSB22 (NSC active and neuronal enhancers) (all these three enhancers are activated by exogenous Ascl1 and inhibited by exogenous Sox2); MSB24 (NSC active enhancer, activated by exogenous Sox2 and inhibited by exogenous Ascl1); MSB11 (NSC active and neuronal enhancer for which no apparent interactions were observed between Ascl1 and Sox2). To limit the number of binding motifs analysed, for enhancers MSB22 and MSB24, I cloned a shorter enhancer element limited to the region where Ascl1 and Sox2 ChIP-seq peaks overlap (Figure 4.1, purple rectangles, see diagrams for MSB 22 and MSB 24). In this way, I limited the mutagenesis analysis to the motifs located in the region where the two TFs overlap and bind in close proximity, according to the ChIP-seq data. The new elements were named MSB22-short and MSB24-short and their

activity was tested and compared to the full-length elements MSB22 and MSB24 (Figures 4.2C and G; 4.3E and H). Enhancer-constructs carrying mutations in each single binding motif were generated. Constructs with more mutated motifs combined in the same elements were also generated for some enhancers.

4.3 Mutations in the sequence of the binding motifs affect enhancer activity

4.3.1 Basal transcriptional activity of the mutant enhancers

Mutant and wild-type enhancers were cotransfected in NS5 cells and their basal transcriptional activity were tested and compared to the activity of the enhancer-less vector (empty vector) by luciferase assay. These experiments should suggest if endogenous *Ascl1* and *Sox2* or other endogenous bHLH and Sox factors expressed in the NS5 cells can induce the activity of enhancers in NSCs through direct binding to the E-boxes and Sox motifs. Indeed *Ascl1* and other bHLH factors bind the same E-boxes, and *Sox2* and other Sox factors bind the same Sox motifs. The activity of the enhancers carrying disrupted motifs should be reduced compared to the activity of the wt sequences if binding and regulation by these endogenous TFs occur.

4.3.1.1 MSB18 Mutants basal activity

MSB18 is both an enhancer active in neural stem cells since it displayed basal luciferase activity in NS5 cells, and a neuronal enhancer induced by overexpression of *Ascl1* (Figure 3.9B, chapter 3). A total of five binding motifs, among E-boxes and Sox motifs, were identified in its sequence (Figure 4.2B). Four of them were located in the

overlapping binding peaks for Ascl1 and Sox2, and only 1 Sox motif was found out of the peak, 85 bp away from the E-box 1. All motifs were mutated individually, generating five mutant elements, named M1 to M5 as shown in the diagram.

Mutation and loss of E-box2 (MSB 18 - M2), E-box3 (MSB 18 - M3), or Sox motif 2 (MSB 18 - M5) strongly reduced the basal activity of these elements compared with the activity of the wt sequence (Figure 4.2A). This result suggests that endogenous TFs binding to these motifs, possibly Ascl1 and Sox2 among them, accounted for the basal activity of the enhancer observed in neural stem cells. Sox motif 1 (MSB 18 - M4) showed a minor reduction, while the luciferase activity of MSB 18 - M1 (E-box1 mutant) did not show any change compared with the WT, suggesting that this motif did not contribute to the enhancer activity in response to endogenous TFs in NS5 cells.

4.3.1.2 MSB22-short Mutants basal activity

MSB22 is also both an enhancer active in neural stem cells since it displayed basal luciferase activity in NS5 cells, and a neuronal enhancer induced by overexpression of Ascl1 (Figure 3.9B, chapter 3). As mentioned earlier, MSB22-short was cloned as shorter element of enhancer MSB22 full-length. Thus, we analysed the input of one E-box and two Sox motifs in the regulation of MSB22-short (Figure 4.2D). Mutation and loss of the E-box (MSB 22 short – M1) or Sox motif 1 (M2) reduced the activity of 2-fold compared with MSB22-short WT (Figure 4.2C), whereas mutation of Sox motif 2 (M3) did not affect the enhancer activity. Also the activity of the mutant elements where mutations of motifs were combined together showed a reduction of the activity compared with the wt sequence, although the combination of mutated motifs did not further affect the reduction as it could be expected (Figure 4.2C). Altogether, this data indicate that endogenous Ascl1 and Sox2 expressed in NS5 cells might activate this

enhancer in neural stem cells through direct binding to the motifs identified, precisely the E-box and Sox motif 1.

4.3.1.3 MSB 4 Mutants basal activity

MSB4 is a neuronal enhancer activated by overexpression of *Ascl1* in NS5 cells (Figure 3.9C, chapter 3). We found one E-box and one Sox motif in its sequence in the region where *Ascl1* and Sox2 ChIP-seq peaks overlap, suggesting DNA binding of the two TFs to the same core region of the enhancer (Figure 4.2F). Mutation and loss of the E-box and Sox motif in MSB4 did not affect the transcriptional activity of the enhancer (Figure 4.2E). This result is consistent with the lack of basal activity showed by the MSB4 enhancer in the neural stem cells, and suggests that binding of endogenous *Ascl1* and Sox factors to this element was not sufficient to induce enhancer activity. Conversely, the combined mutation of the two binding motifs in mutant element M3 increased the enhancer activity compared with the wt element, suggesting that factors binding to the E-box and Sox motif may actually repress this enhancer (Figure 4.2E).

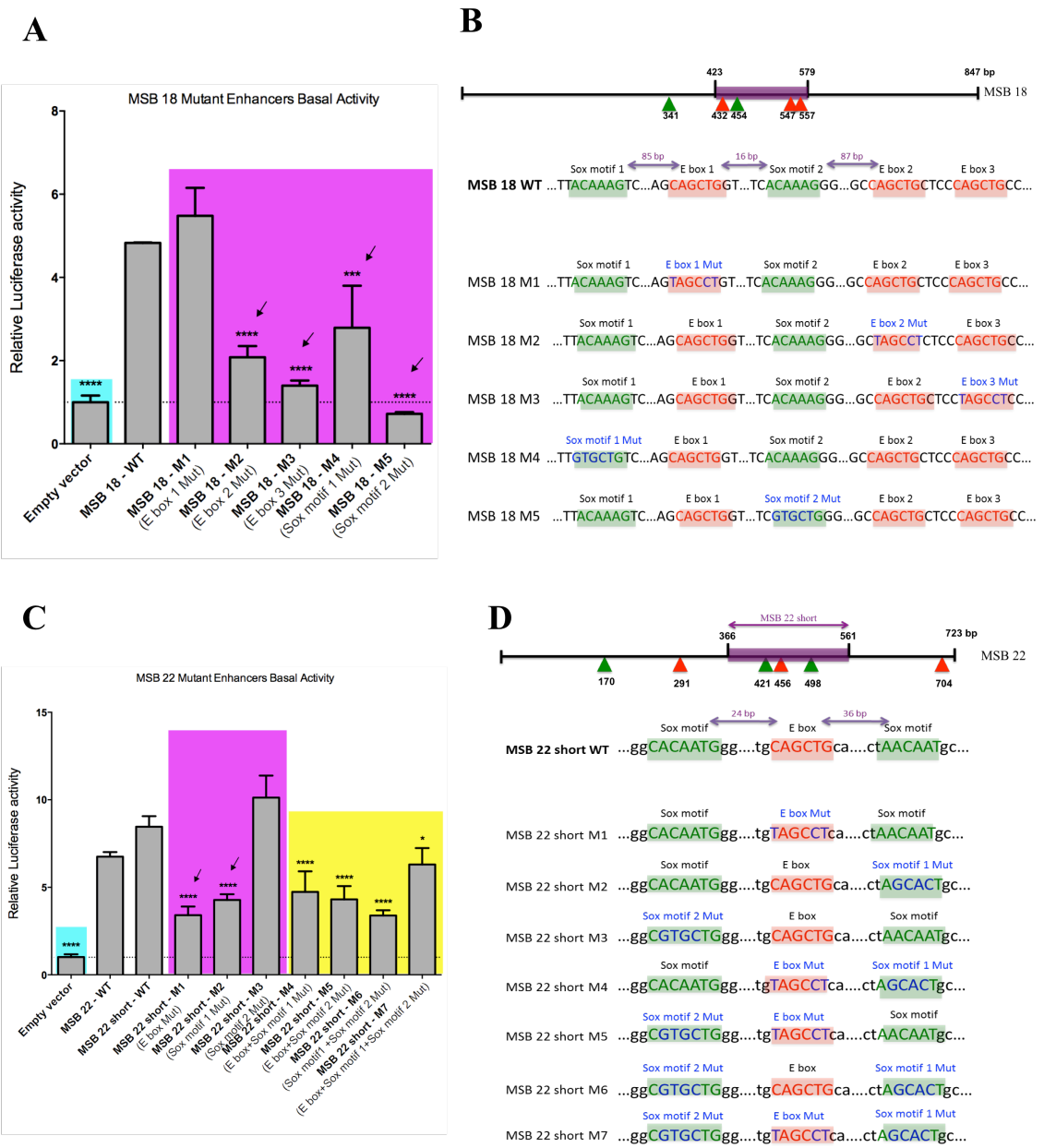
4.3.1.4 MSB24-short and MSB11 Mutants basal activity

MSB24 is an enhancer active in neural stem cells, due to the basal luciferase activity shown in NS5 cells (Figure 3.9A, chapter 3). MSB11 is both an enhancer active in neural stem cells, and a neuronal enhancer activated by overexpression of *Ascl1* in NS5 cells (Figure 3.9B, chapter 3). The binding motifs identified in these enhancers are shown in the diagrams in Figure 4.2H and J, respectively. The basal activity of MSB24-short is slightly increased compared with the activity of the full-length enhancer, suggesting that binding motifs located in the sequence outside the overlapping *Ascl1* and Sox2 ChIP-seq peaks could repress the activity of the full-length element (Figure

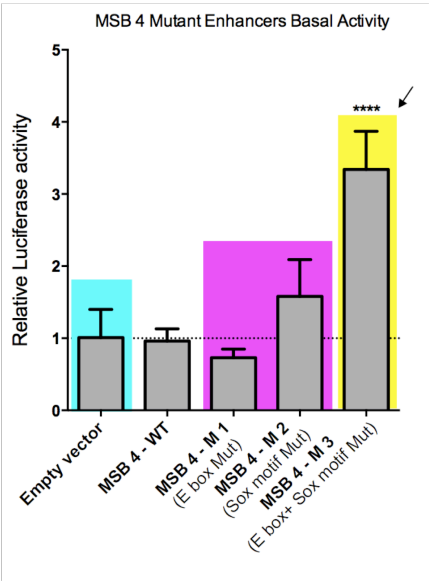
4.2G). Mutations and loss of the binding motifs in MSB24-short and MSB11 (Figure 4.2G and I) did not affect the basal activity of the enhancers or only showed a minor reduction, respectively, suggesting that the binding motifs identified and mutated did not give great contribution to the transcriptional activity of these elements and to their regulation by endogenous TFs in neural stem cells (Figure 4.2G and I).

In conclusion, the results of the mutagenesis analysis showed that loss of the E-box and Sox motifs affected the basal transcriptional activity of some enhancers, such as MSB18 and MSB22-short. However, other enhancers were not affected when their BMs were lost, such as MSB4, MSB24-short, or only showed minor reduction, as for MSB11. Furthermore, results also distinguished between different BMs within the same enhancer, with some involved in the regulation of the element, as showed by a change in the activity of their mutants, and others excluded. Overall, basal luciferase activity of some mutants could suggest an input of endogenous *Ascl1* and *Sox2* in the regulation of some enhancers, as for MSB18 and MSB22-short. However, for the other elements, MSB24-short, MSB11, and MSB4, loss of the E-box and Sox motif did not greatly affect the basal activity of the mutant enhancers. This result might indicate that the BMs identified and mutated, and therefore endogenous *Ascl1* and *Sox2* among the TFs possibly binding these sites, did not account for the basal transcriptional activity of the enhancers in neural stem cells, as it seems the case for MSB24-short. This element was active in the neural stem cell context, as shown in the previous chapter, and was activated by exogenous *Sox2*. Therefore, the results of the mutagenesis suggest that other binding motifs here not identified are involved in the regulation of the basal activity in NS5 cells. It is also possible that TFs, such as *Sox2*, bind other less canonical motifs and/or regulate the element via indirect DNA binding. In the case of MSB4 however the result is compatible with the lack of basal activity showed by the wt

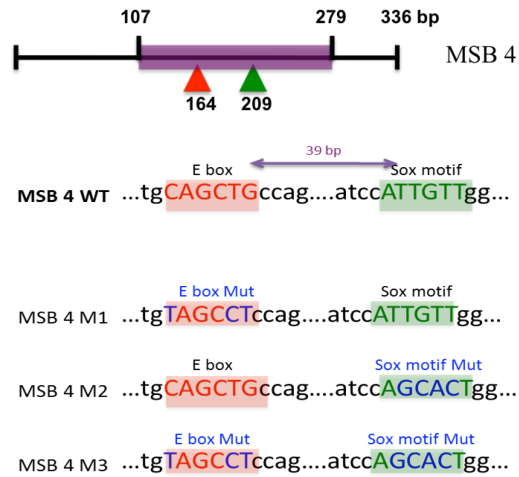
enhancer in neural stem cells. MSB4 is a neuronal enhancer activated by exogenous *Ascl1* in differentiating neurons. However, this element is not active in the neural stem cell context, as shown in the previous chapter.



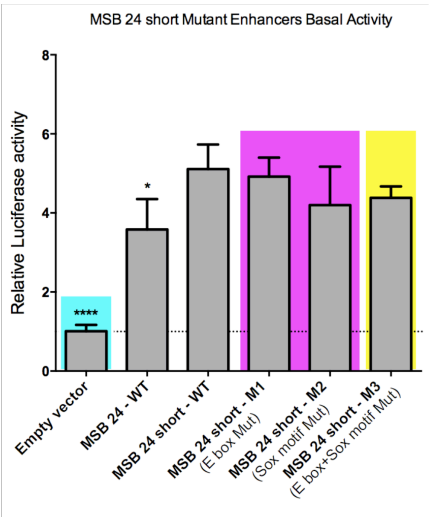
E



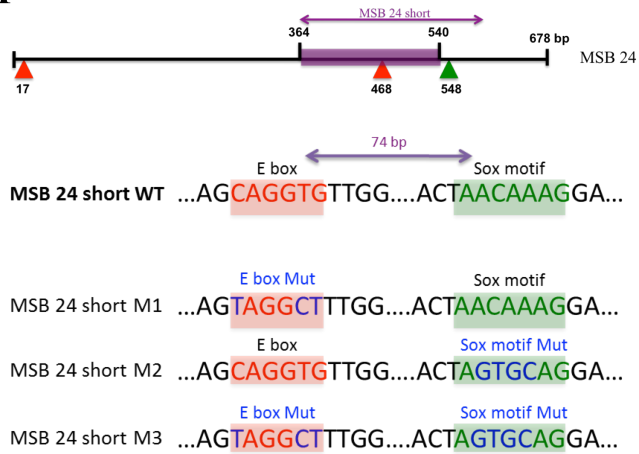
F



G



H



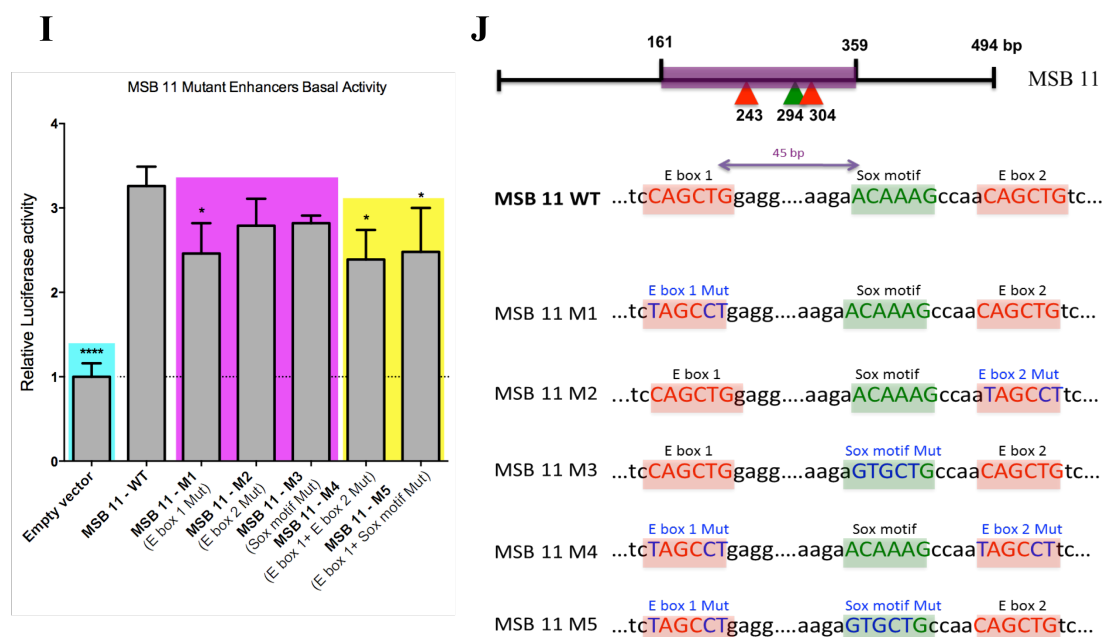


Figure 4.2. Mutagenesis of the E-box and Sox motifs affects basal luciferase activity of the enhancers.

Wt and mutant enhancers carrying disrupted E-box and Sox motifs were cotransfected in NS5 cells. Luciferase activity was measured 24 hours after transfection. Luciferase activity was normalised to the activity of the empty vector (on blue background in the pictures). Data are presented as the mean \pm SD of triplicate assays. Mutant enhancers carrying mutation of a single motif are shown on purple background and mutant enhancers carrying combined mutations of more than one motifs are shown on yellow background. Panels B, D, F, H, and J represent diagram of each enhancer as in Figure 4.1 and the sequence of the motifs within each element. Mutated motifs are labeled in blue at the top of the sequence and mutated nucleotides are shown in red. Distances between adjacent motifs are shown as purple arrows at the top of the wt element. (A) In enhancer MSB18, mutations of the E-box 2, E-box 3, and Sox motif 2 strongly reduced the basal activity of the mutant elements compared to the wt (as indicated by the black arrows). Mutation of the motifs in mutant elements M4 and M1 show a minor reduction or no change in their activity compared with the wt element, respectively. (C) In enhancer MSB22-short, shorter element of enhancer MSB22 full-length (corresponding to the purple rectangle in panel D, see text for full explanation), only mutation of the E-box (mutant element M1) and Sox motif 1 (element M2) reduced the basal activity of the mutant compared with the wt short enhancer. Also mutant elements M4 to M7 with combined mutations of motifs (on yellow background) exhibited reduced basal activity compared with the wt, but combination of mutated motifs did not further affect the reduction. (E) Mutation of the E-box and Sox motif did not change the basal activity of mutant elements M1 and M2 compared with the wt enhancer MSB4. Combined mutation of the two binding motifs in mutant element M3 increased the enhancer activity compared with the wt element (as indicated by the arrow). (G) and (I) Mutations of the E-box and Sox motifs in enhancers MSB24-short and MSB11 did not change or only slightly reduced the basal activity of these elements. Basal activity was not changed or only slightly reduced also when mutations of motifs were combined in the same mutant elements, on yellow background.

4.3.2 Mutagenesis of the E-box and Sox motifs affects the regulation by exogenous Ascl1 and Sox factors in the neural enhancers

Mutant and wt enhancers were cotransfected in NS5 cells together with expression vectors for Ascl1 and Sox2 overexpressed either alone or simultaneously. After 24 hours, cells were harvested and the luciferase activity of the mutant and wt enhancers were measured and compared. The purpose of this experiment was to understand if mutations of the binding motifs could affect the regulation by exogenous Ascl1 and Sox2, suggesting that the two TFs regulate the enhancers through direct DNA binding to their motifs. For each enhancer, the luciferase activity was expressed as relative luciferase activity compared to the activity of the same construct after overexpression of a GFP expression vector used as control.

4.3.2.1 The neuronal enhancer MSB4

MSB4 is a neuronal enhancer strongly activated by exogenous Ascl1 overexpressed in NS5 cells. Overexpression of Sox2 counteracts Ascl1-induced activity of this element (Figure 3.8A, chapter 3). The binding motifs identified in this enhancer and previously described are shown in Figure 4.3B. Point mutations of the E-box, known to disrupt the interaction of Ascl1 with its consensus (Castro et al., 2006), abolished the activation of this enhancer by Ascl1 when the TF was overexpressed either alone or together with Sox2, although activation still persisted when the two TFs were overexpressed together (Figure 4.3A). This result suggests that exogenous Ascl1 activates this element through direct DNA binding to the E box. Point mutations of the Sox binding motif, described to disrupt the interaction of Sox2 with its consensus (Tanaka et al., 2004), did not affect the regulation of the mutant M2 by exogenous Sox2. Unexpectedly, the loss of the Sox motif disrupted the activation of the enhancer by exogenous Ascl1. The

activity of the mutant elements M1 and M2, the E-box mutant and the Sox motif mutant elements respectively, were very similar after overexpression of Ascl1 (Figure 4.3A). The activation of the mutant M2 was compromised also when Ascl1 and Sox2 were overexpressed together. The result of this experiment did not confirm binding and requirement of Sox motif for the exogenous Sox2 regulation of the neuronal enhancer MSB4. However, it is worth noting that Sox2 did not show regulation of this enhancer, as previously described. For clearer conclusions on the mechanism of regulation by Sox2, experiments like EMSA (Electrophoretic Mobility Shift Assay) can prove or exclude direct binding of Sox2 to its motif.

In conclusion, the most important result from the experiments of mutagenesis is the requirement of the Sox motif for Ascl1-induced activation of the neuronal enhancer MSB4. This suggests that Ascl1 might interact with a Sox factor binding to the identified Sox motif in the activation of the neuronal enhancer.

4.3.2.2 The NSC-active and neuronal enhancer MSB18

MSB18 is both an enhancer active in neural stem cells showing basal luciferase activity in NS5 cells, and also a neuronal enhancer activated by exogenous Ascl1. Sox2 counteracts Ascl1-induced activation of this element (Figures 3.8A and 3.9B, chapter 3). The five BMs identified and mutated to generate the MSB 18 mutant enhancers are shown in Figure 4.3D. Mutagenesis of enhancer MSB18 shows that different binding motifs give different contribution to the regulation of the enhancer. More precisely, only mutant M1 (E-box1 mutant) showed a reduction of activity when the E-box was mutated. When the E-box 1 was mutated, mutant M1 lost the activation by exogenous Ascl1 (Figure 4.3C). Loss of the Sox motif in the mutant construct M4 (Sox motif 1 mutant) did not affect regulation of the enhancer by exogenous Sox2 (Figure 4.3C). The

stem cell associated TF Sox2 did not regulate this neuronal enhancer, consistently with previous result discussed in paragraph 3.5.2 (Figure 3.8A, chapter 3). Conversely, mutants MSB18 M2, M3, and M5 (E-box 2, E-box 3, and Sox motif 2 mutant elements, respectively) showed a robust increase of activity by exogenous Ascl1 compared with the wt element, when the TF was overexpressed either alone, or together with Sox2. M5 mutant was also strongly activated by exogenous Sox2 compared with the wt element (Figure 4.3C). This result suggests that the Ascl1-induced activation of the wt enhancer might be repressed by other TFs, possibly other bHLH proteins, binding to the other E-boxes 2 and 3, or by a Sox factor binding to the Sox motif 2. The Sox factor not identified here binding to Sox motif 2 represses also the activity of Sox2 in the regulation of enhancer MSB18. The result distinguishes among the E-boxes bound by Ascl1, which might bind E-box 1 rather than the others, according to this experiment.

In conclusion, results of this experiment suggest that activation of enhancer MSB18 occur through direct DNA binding of Ascl1 to the E-box1. It is possible that other bHLH TFs binding E-box 2 and 3 or a Sox factor binding Sox motif 2 repress Ascl1 and Sox2 in the activation of the wt element.

4.3.2.3 The NSC-active and neuronal enhancer MSB 22-short

MSB22 has been classified as both a neural stem cell active enhancer, and also a neuronal enhancer activated by overexpression of Ascl1 in NS5 cells. Sox2 strongly counteracts Ascl1-induced activation of this element (Figures 3.8A and 3.9B, chapter 3). MSB22-short with its motifs is displayed in Figure 4.3G. The regulation of MSB22-short by exogenous Ascl1 and Sox2 was tested and compared to the activity of MSB22 full-length (Figure 4.3E). MSB22-short still exhibited characteristics of a neuronal enhancer, induced by exogenous Ascl1. However, Ascl1-induced activation of the short

element was lower than the full-length element. This suggests that Ascl1 might also bind the other two E-boxes of the full-length element with an additive effect in the regulation of the entire element (Figure 4.3E and G). Activation of the short element by exogenous Sox2 was increased compared to the full-length element. This is probably because the short element has two Sox motifs and only one E-box. Ascl1 and Sox2 still appeared to counteract each other in the regulation of MSB22-short, despite the different number of motifs present in the sequence. Mutation of the E-box abolished activation of enhancer MSB22-short by exogenous Ascl1. This result suggests that Ascl1 regulates the activity of the enhancer through direct DNA binding to its consensus motif (Figure 4.3F). Mutation of either Sox motif 1 or Sox motif 2 did not affect the regulation of the enhancer by Sox2. This was not affected also when the two Sox motifs were mutated together in the same mutant element M6 (Figure 4.3F). This result suggests that Sox2 might regulate enhancer MSB22-short without direct binding to its consensus motifs here identified.

Interestingly, activation of MSB-22 short enhancer by exogenous Ascl1 was highly increased in mutant element M6 where the two Sox motifs were mutated together compared with the activity of the wt sequence (Figure 4.3F). This result suggests that other Sox factors not identified here might bind to these motifs and repress Ascl1 activation of the enhancer. A similar result was observed for Sox motif 2 in enhancer MSB18.

In conclusion, results from this experiment show that Ascl1 activates the neuronal enhancer MSB22 through direct binding of the E-box identified. However, results could not confirm binding of Sox2 to its motifs, suggesting a different mechanism of regulation by this TF. The mutagenesis analysis of enhancer MSB22-short suggest that also other Sox factors could be involved in the regulation of this element and they might

counteract Ascl1 activity. Results reveal a complex and context-specific interaction between Ascl1 and the Sox factors in the regulation of this enhancer.

4.3.2.4 The NSC-active enhancer MSB24-short

MSB24 is an enhancer active in neural stem cells as shown by basal luciferase activity in NS5 cells (Figure 3.9A, chapter 3). This enhancer is activated by overexpression of Sox2 in NS5 cells. Ascl1 counteracts Sox2-induced activation of the element when the two TFs are overexpressed together (Figure 3.10A). Diagram in Figure 4.3I represents MSB24 enhancer and the binding motifs identified and mutated. MSB24-short was cloned as shorter element of MSB24. The activity of MSB24-short and MSB24 full-length were tested and compared in NS5 cells. The two elements display similar pattern of activity (Figure 4.3H). Activation of the short element by exogenous Sox2 is increased compared with the full-length element. This suggests that a bHLH factor binding to the second E-box present in the full-length enhancer contributes to the repression of exogenous Sox2 activity (Figure 4.3I). Mutation disrupting the E-box did not affect regulation of MSB24-short M1 mutant enhancer by exogenous Ascl1. Mutation of the Sox motif exhibited a minor reduction of the activation of the enhancer by exogenous Sox2 compared with the WT element (Figure 4.3H, mutant M2 vs wt element).

In conclusion, these results might suggest direct binding of Sox2 to its motif in the regulation of the NSC active enhancer MSB24-short, although reduction of the activation by exogenous Sox2 is low. EMSA assay can confirm direct DNA binding of this TF to its motif and gives better insight in the mechanism of regulation.

4.3.2.5 The NSC-active and neuronal enhancer MSB 11

MSB11 has been classified as both a neural stem cell active and also a neuronal enhancer (Figure 3.9B). Diagram in Figure 4.3K represents enhancer MSB11 and the binding motifs identified and mutated. Overexpression of *Ascl1* or *Sox2* shows only minimal effect in the regulation of this enhancer (Figure 4.3J). The activity of the wt element was minimally increased when the two TFs were overexpressed simultaneously in NS5 cells. This suggests additive effects of the two regulators, which might act independently. Mutation of the E-box 1 in mutant MSB11-M1 showed a minor reduction by exogenous *Ascl1* compared with the wt element suggesting that this TF might directly bind to its motif for the regulation of this enhancer. Overall, variations of the activity of the mutant constructs compared with the wt element are only minimal and it is hard to draw conclusions on the interaction between *Ascl1* and *Sox2* in the regulation of this enhancer.

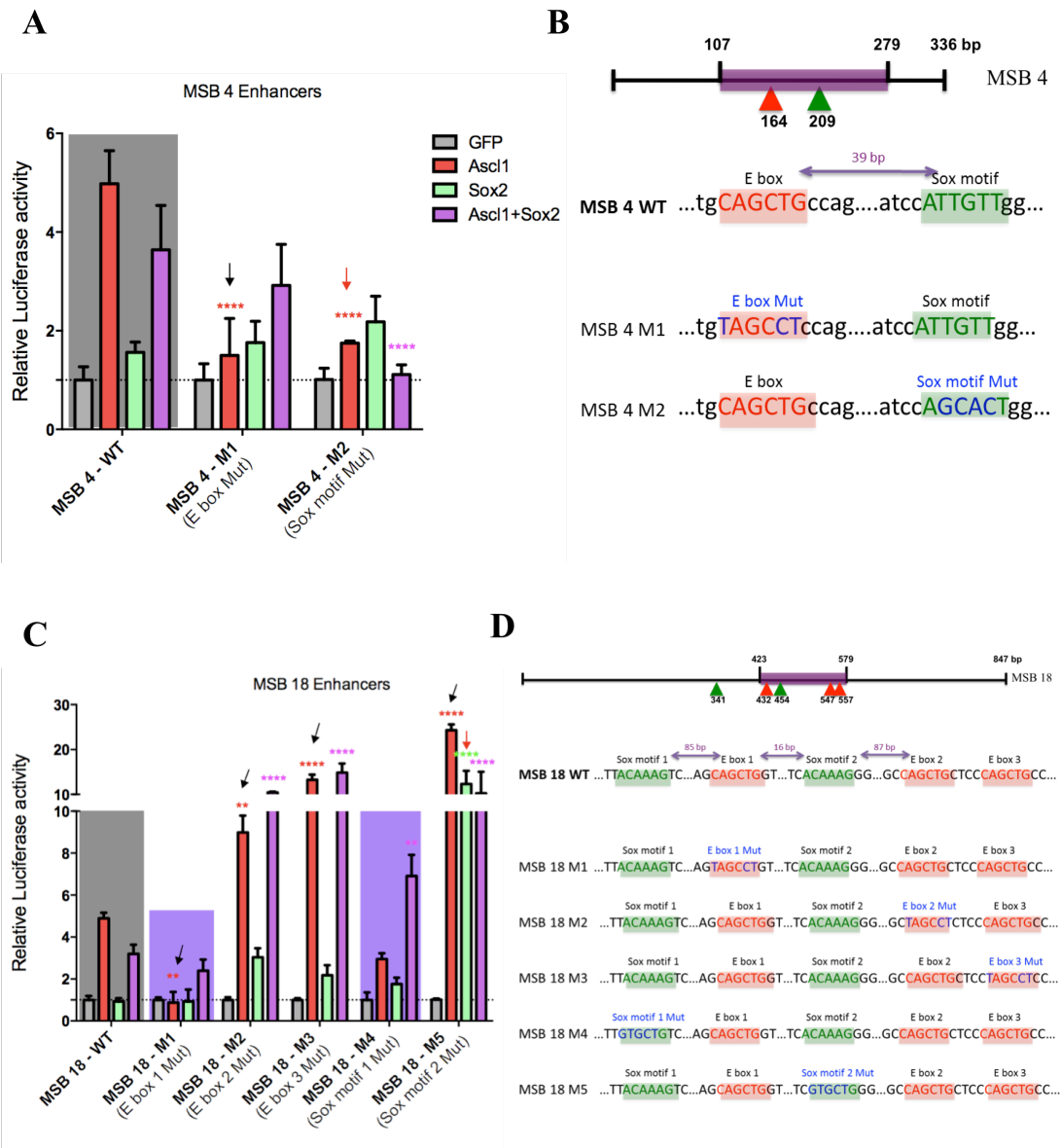
In conclusion, the mutagenesis of the TF binding motifs of the neural enhancers gave in depth insight in the mechanisms of transcriptional regulation, in particular for the neuronal enhancers MSB4, MSB18 and MSB22. Results give evidence that *Ascl1* activates these elements through direct DNA binding to the E-boxes here identified and mutated. Results also distinguish among binding motifs identified within the same regulatory element. For instance, *Ascl1* appears to bind specifically E-box 1 but not E-box 2 and E-box 3 in enhancer MSB 18.

Mechanisms of regulation of the neural enhancers by exogenous *Sox2* appear to be more elusive, since mutations of the *Sox* motifs did not suggest direct binding of this TF for most of the elements mutated, with the exception of MSB24 enhancer. In this case, mutation of the *Sox* motif caused a minor reduction of the activation by exogenous

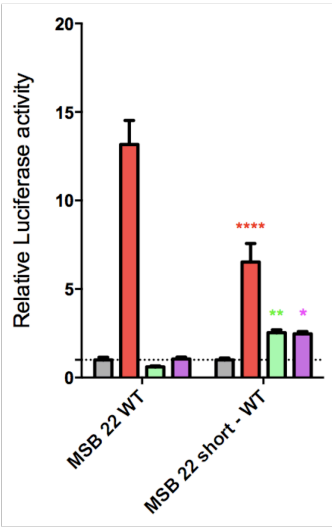
Sox2. Direct DNA binding of Sox2 to enhancer MSB24 needs to be further investigated. It is possible that Sox2 regulates the neural enhancers without direct DNA binding to the identified motifs. However, it is worth noticing that Sox2 acts as a weak regulator of the transcriptional activity of all enhancers identified in this study. In this perspective, the lack of effect of disrupted Sox motifs in the regulation of these enhancers is consistent with a poor function of Sox2 as transcriptional activator of these elements. Binding of Sox2 might be non-functional in the transcriptional activity and output of these enhancers. Therefore, mutation of the Sox motifs disrupting the binding might equally not show any effect in the regulation of the mutated enhancers.

Moreover, mutagenesis analysis of the Sox motifs shows different roles for different Sox motifs identified, suggesting that different Sox factors might be recruited and involved in the regulation of the neural enhancers identified in this study. For instance, another Sox factor rather than Sox2 might bind to Sox motif 2 and repress the activity of enhancer MSB18. Therefore, other experiments are required to gain insight on the recruitment of the Sox factors and their regulation of the neural enhancers identified in this study.

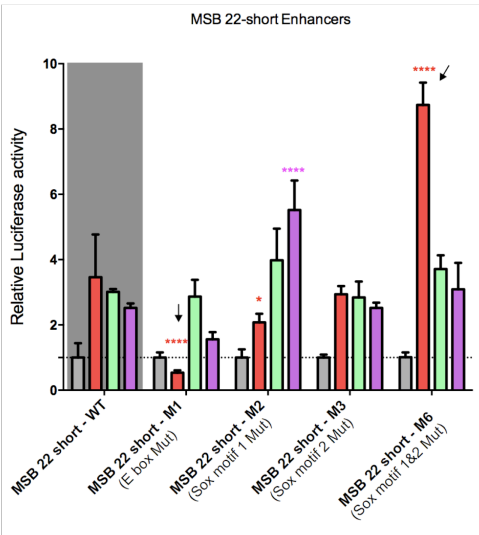
Finally, the most interesting result emerging from the mutagenesis analysis is the requirement of a Sox motif for Ascl1-induced activation of the neuronal enhancer MSB4. This result suggests that a Sox factor binding to this motif might enhance recruitment or activity of Ascl1 in the regulation of MSB4. It is also possible that protein-protein interactions are involved between Ascl1 and a Sox factor binding to the identified motif in the regulation of this enhancer. This mechanism needs to be further dissected.



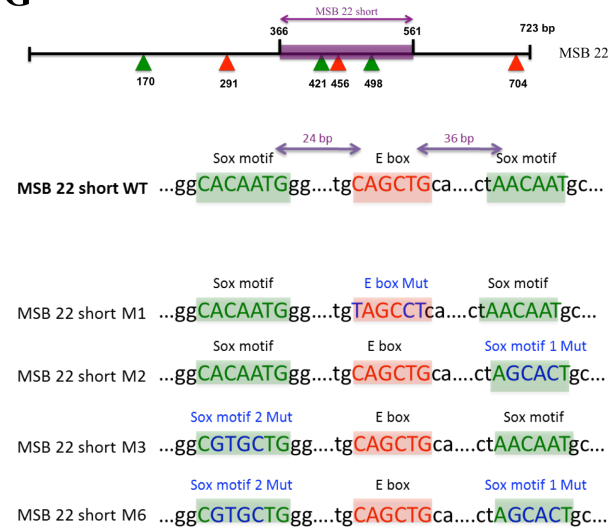
E



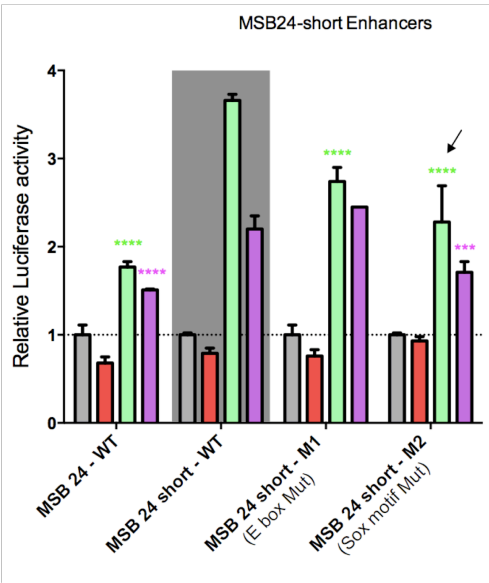
F



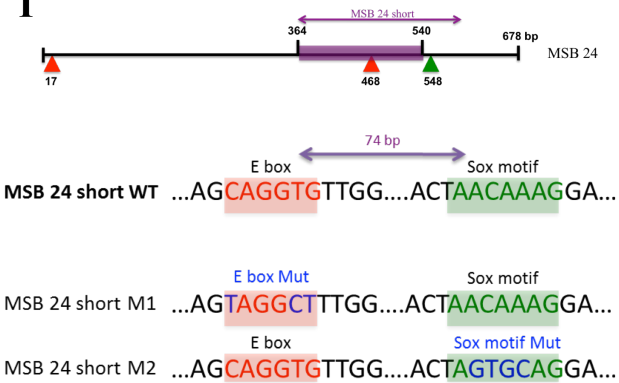
G



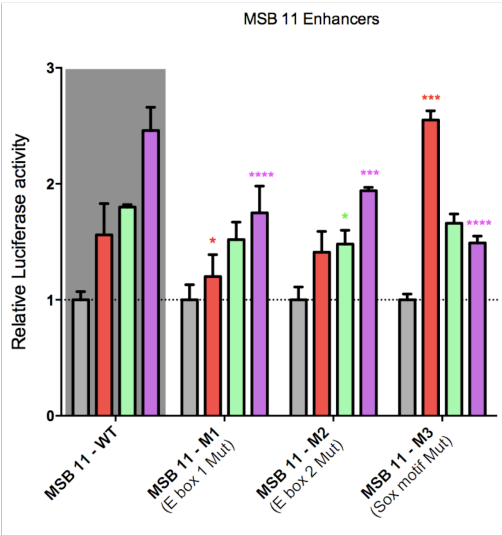
H



I



J



K

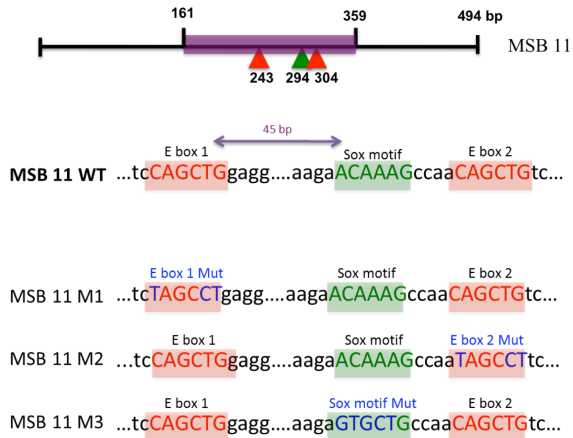


Figure 4.3. Mutagenesis of the E-box and Sox motifs affect the regulation by exogenous Ascl1 and Sox factors in the neural enhancers.

Mutant and wt enhancers were cotransfected in NS5 cells together with expression vectors for Ascl1 and Sox2 overexpressed either alone or simultaneously (as shown in the legend of the graphs). Luciferase activity was measured after 24 h. For each enhancer, luciferase activity was normalised to the activity of the same enhancer after overexpression of GFP as control. For each enhancer, luciferase activity of the wt element is shown on grey background in the graphs. Data are presented as the mean \pm SD of triplicate assays. **Panels B, D, G, I, K J** represent diagram of each enhancer as in Figure 4.1 and the sequence of the motifs within each element, as explained in Figure 4.2. **(A)** Mutation of the E-box abolished the activation of enhancer MSB4 by exogenous Ascl1 (as shown by the black arrow in the graph). Mutation of the Sox motif did not change the activity by exogenous Sox2. Mutation of the Sox motif also abolished activation of the enhancer by exogenous Ascl1 (mutant element M2 vs wt) as shown by the red arrow in the graph. The experiment shown for enhancer MSB4 is representative of three independent biological replicates. **(C)** Mutagenesis of enhancer MSB18 shows that different binding motifs give different contribution to the regulation of the enhancer. Mutation of the E-box 1 abolished Ascl1-induced activation of this neuronal enhancer (mutant element M1 vs wt, as indicated by the arrow in the graph). Mutation of the Sox motif 1 did not change the activity by exogenous Sox2, which did not show activation of the enhancer. Mutation of the E-box 2, E-box 3 and Sox motif 2 (mutant elements M2, M3, and M5 vs wt element) increased strongly the activation of the enhancer by exogenous Ascl1 and also by exogenous Sox2 in mutant M5 (as indicated by the three black arrows in the graph) when Ascl1 was overexpressed alone or together with Sox2, respectively. **(E)** Luciferase activity of neuronal enhancer MSB22-short compared with MSB22 full-length after overexpression of Ascl1 and Sox2 either alone or together. Exogenous Sox2 still inhibited the activation of the enhancer by exogenous Ascl1 in both elements. MSB22-short exhibited a reduced activation by exogenous Ascl1 and an increased activation by exogenous Sox2 compared with the full-length element. **(F)** In enhancer MSB22-short, mutation of the E-box abolished activation of the element by exogenous Ascl1 (as shown by the black arrow) (mutant element M1 vs wt element on grey background). Mutation of the Sox motif did not affect regulation by exogenous Sox2, also when the two Sox motifs were mutated together in the same element (mutant element M6, and also M2 and M3 vs wt). Mutation of the two Sox motifs in mutant element M6 increased the activation of the enhancer by exogenous Ascl1. The experiment shown for enhancer MSB22-short is representative of two independent biological replicates. **(H)** Enhancer MSB24-short wt shows an increased activation by exogenous Sox2 compared with the full-length wt element. Mutation of the E-box and Sox motif did not change the values of activity by exogenous Ascl1, which doesn't regulate this NSC active enhancer, as shown in the wt element (either full-length or short). Mutation of the Sox motif reduced minimally the activation of the enhancer by exogenous Sox2. **(J)** For enhancer MSB11, overexpression of Ascl1 or Sox2 shows only minimal effect in the regulation. The activity of the wt element was minimally increased when the two TFs were overexpressed simultaneously in NS5 cells.

4.4.1 SoxC factors, Sox4 and Sox11, synergise with Ascl1 in the regulation of the neuronal enhancers MSB4, MSB18 and MSB22

Mutagenesis analysis of the binding motif of the neuronal enhancers MSB4 showed the requirement of the Sox motif for the Ascl1-induced activation of this element. However, given that this enhancer was not induced by Sox2 and since Sox2 appeared to counteract Ascl1 in the activation, we hypothesised that other Sox factors rather than Sox2 could bind to the same Sox motif and synergise with Ascl1 in the activation of this neuronal enhancer. This hypothesis was supported by the fact that Sox factors tend to

recognise and bind to the same consensus sequences (Lefebvre et al., 2007). Thus, the Sox motifs identified were not exclusively Sox2 binding motifs, but could have been targeted by different Sox factors. RNA-seq data generated in our lab showed that other Sox factors were transcribed in the NS5 cells, including Sox4, Sox11, and also Sox 8 (see next paragraph), with Sox4 being the highest transcribed (Martynoga et al., 2013). However, at the present we do not have data about expression at protein levels for these TFs in the NS5 cells. Also lab data from arrays after *Ascl1* Gain of Function (GoF) showed induction of Sox4, Sox11, and Sox8, while showing repression of Sox2 (Martynoga, unpublished data). Moreover, the literature shows that Sox4 and Sox11, members of the SoxC family of TFs, are induced by proneural proteins, and have a critical role in the establishment of neuronal properties (Bergsland et al., 2006).

Taken together, our results and the literature prompted us to investigate if SoxC factors, in particular Sox4 and Sox11, could activate the neuronal enhancers and if they could synergise with *Ascl1* in their regulation. We carried on with luciferase assays to assess the transcriptional activation of the neuronal enhancers MSB4, MSB18, and MSB22 by Sox4 and Sox11. As previously described, the enhancers were transfected in NS5 cells with expression vectors for *Ascl1*, and Sox4/Sox11 overexpressed either alone or simultaneously. Cells were harvested 24 hours after transfection and the luciferase activity was measured. As shown in Figure 4.4A and B, although the neuronal enhancers analysed were not induced by Sox4 and Sox11 when these TFs were overexpressed alone, all neuronal enhancers MSB4, MSB22 full-length and the short element, and MSB18 were strongly induced by *Ascl1* and Sox4/Sox11 when the two TFs were overexpressed together. Therefore, we proved in this experiment that SoxC factors, Sox4 or Sox11, strongly synergise with *Ascl1* in the activation of the neuronal enhancers identified in this study (black arrows, figure 4.4A and B).

4.4.2 Dual function of Sox8, member of the Sox E family, in the regulation of the neuronal enhancers MSB 4, MSB 18 and MSB 22

RNA-seq data and microarray data from Ben Martynoga mentioned above showed that also Sox8, a member of the SoxE family of TFs, is expressed in NS5 cells and is induced by Ascl1 GoF. Therefore, we also tested the activity of Sox8 in the regulation of the neuronal enhancers in luciferase assays. As showed in Figure 4.4C, Sox8 did not activate enhancers MSB4 and MSB18 when it was overexpressed alone, but strongly synergised with Ascl1 to activate these neuronal enhancers when the two TFs were overexpressed simultaneously (black arrows, Figure 4.4C). Strikingly, the same TF inhibited enhancer MSB22, either its full-length element or its short version, by counteracting Ascl1 activation of this neuronal enhancer when the two TFs were overexpressed together. MSB22 full-length and MSB22-short were indeed strongly repressed when Ascl1 and Sox8 were overexpressed together in NS5 cells, with values as low as 0.2 (red arrows, Figure 4.4 C).

In conclusion, these results show a dual role for the TF Sox8 as a transcriptional regulator of neuronal enhancer, capable of synergising with Ascl1 in the regulation of MSB4 and MSB18, and to strongly repress it in the regulation of MSB22 and MSB22-short. In the case of enhancer MSB22, Sox8 displayed a similar tendency as Sox2 to counteract Ascl1 in the regulation of the identified neuronal enhancers.

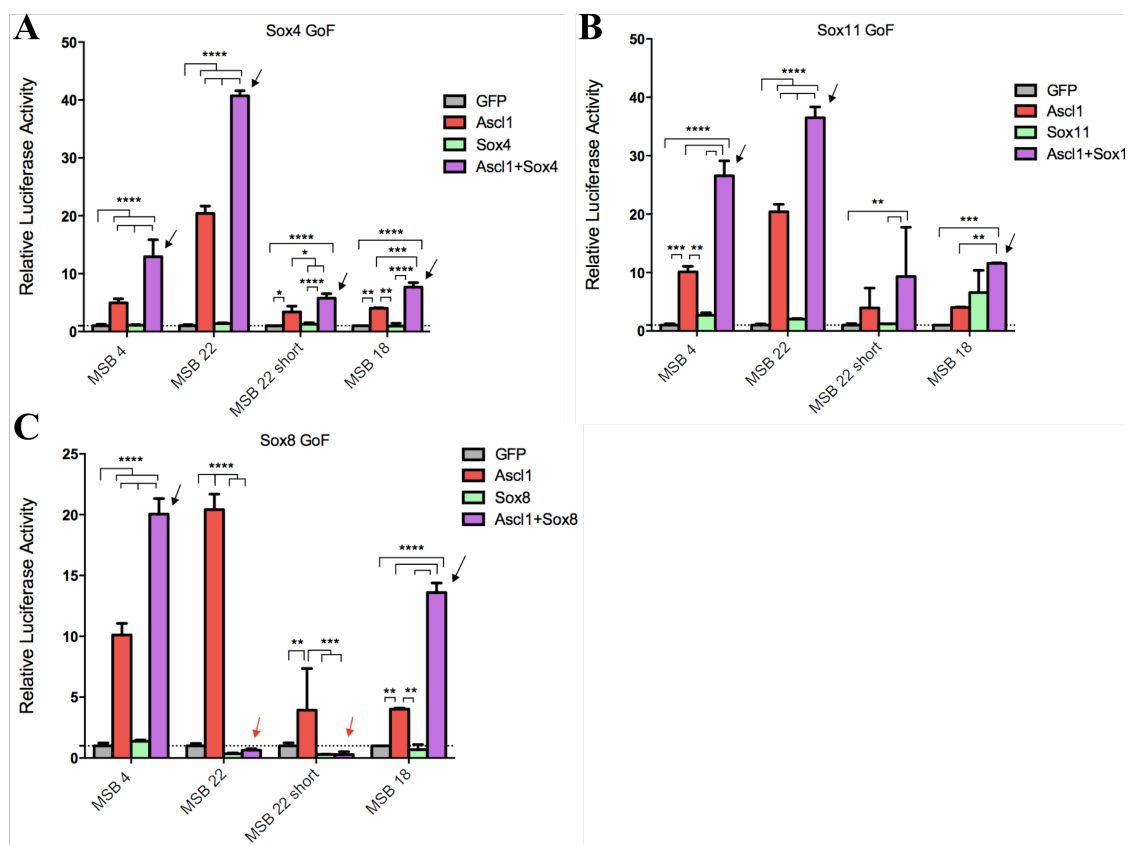


Figure 4.4. Interplay between SoxC or Sox8 factors and Ascl1 in the regulation of the neuronal enhancers.

Neuronal enhancers were transfected in NS5 cells with expression vectors for Ascl1, Sox4/Sox11 or Sox8 overexpressed either alone or simultaneously (panel A and B for Sox4 or Sox11 overexpression, respectively; panel C for Sox8 overexpression). Luciferase activity was measured 24 h after transfection. All neuronal enhancers tested (MSB4, MSB22 full-length and short element, and MSB18) were activated by exogenous Ascl1 (A, B, and C). Overexpression of Sox4 and Sox11 did not activate the enhancers (A and B). All neuronal enhancers were synergistically activated by simultaneous overexpression of Sox4 or Sox11, as indicated by the black arrows (in A and B). Overexpression of Sox8 alone did not activate the neuronal enhancers (C). Enhancers MSB4 and MSB18 were synergistically activated by simultaneous overexpression of Ascl1 and Sox8 (black arrows in C) while Sox8 strongly inhibited Ascl1-induced activation of enhancer MSB22 full-length and short (red arrows in C). Data are presented as the mean \pm SD of triplicate assays. The experiment shown is representative of two independent biological replicates for all enhancers but MSB18.

4.5 Mutations of E-box and Sox motifs affect the synergy of Ascl1 with SoxC factors, or Sox8 in the activation of neuronal enhancers

Once we demonstrated the synergy between Ascl1 and Sox4/11, and, in some cases, Sox8 in the regulation of the neuronal enhancers, we wanted to assess if mutation of the binding motifs would have affected this synergy. This would suggest recruitment of Sox factors (Sox4, Sox11, and Sox8) to their binding motifs and the mechanisms of interaction with Ascl1 recruited on the E box. With this purpose, we tested the activity of the same mutant enhancers described in paragraph 4.3.1 after overexpression of Ascl1 and Sox4/Sox8/Sox11. Mutant enhancers and their wt elements were cotransfected in NS5 cells together with expression vectors for Ascl1 and Sox4/Sox11/Sox8 overexpressed alone or simultaneously.

4.5.1 Regulation of MSB4 mutant enhancers by Ascl1 and Sox factors

MSB4 is a neuronal enhancer activated by overexpression of Ascl1 in NS5 cells. As demonstrated in previous experiments (paragraph 4.4.1, Figure 4.4A, B, and C), overexpression of the Sox factors Sox4/Sox11 or Sox8 in NS5 cells do not activate this enhancer. MSB4 is synergistically activated by simultaneous overexpression of Ascl1 and Sox4/Sox11 or Sox8 in NS5 cells (Figures 4.5A,B, and C, MSB4 wt element). Mutation of the E-box abolished the activation of the enhancer by Ascl1, as previously demonstrated, and also the synergy between Ascl1 and Sox4/Sox11 or Sox8, as expected (Figure 4.5A, B, and C, mutant element M1). Mutation of the Sox motif abolished Ascl1-induced activation of the enhancer, as previously demonstrated, and did not affect activity by any of the Sox factors tested, Sox4, Sox11, or Sox8. Mutation

of the Sox motif also strongly reduced the synergistic activation of the enhancer by simultaneous overexpression of *Ascl1* and all Sox factors tested in these experiments (Figures 4.5A, B, and C, mutant element M2). These results cannot suggest if the Sox factors regulate the enhancer through direct DNA binding to the Sox motif here identified. Different experiments, such as EMSA, can confirm or exclude this possibility. These results are very similar to those obtained for Sox2 and discussed in the previous paragraphs. Thus, mechanisms of regulation of the neuronal enhancer MSB4 by all Sox factors analysed in this study need to be further dissected. The synergistic activation of MSB4 enhancer by *Ascl1* and the Sox4/Sox11 or Sox8 is lost either with mutation of the E-box or with mutation of the Sox motif. However, in the second case, it is not possible to distinguish if the synergy is lost as a consequence of the effect of the mutation of the Sox motif on the activity of the exogenous *Ascl1*, or as an effect of the mutation of the Sox motif on the binding of the Sox factors. Indeed, Sox factors might bind to the Sox motif and activate the enhancer in synergy with *Ascl1* whereas they do not exhibit transcriptional activity alone. It is known that Sox factors alone do not appear to be functional and they only exert transcriptional activation of a bound regulatory region in combination with a partner factor (Yuan et al., 1995, Kamachi et al., 2001, Kondoh and Kamachi, 2010).

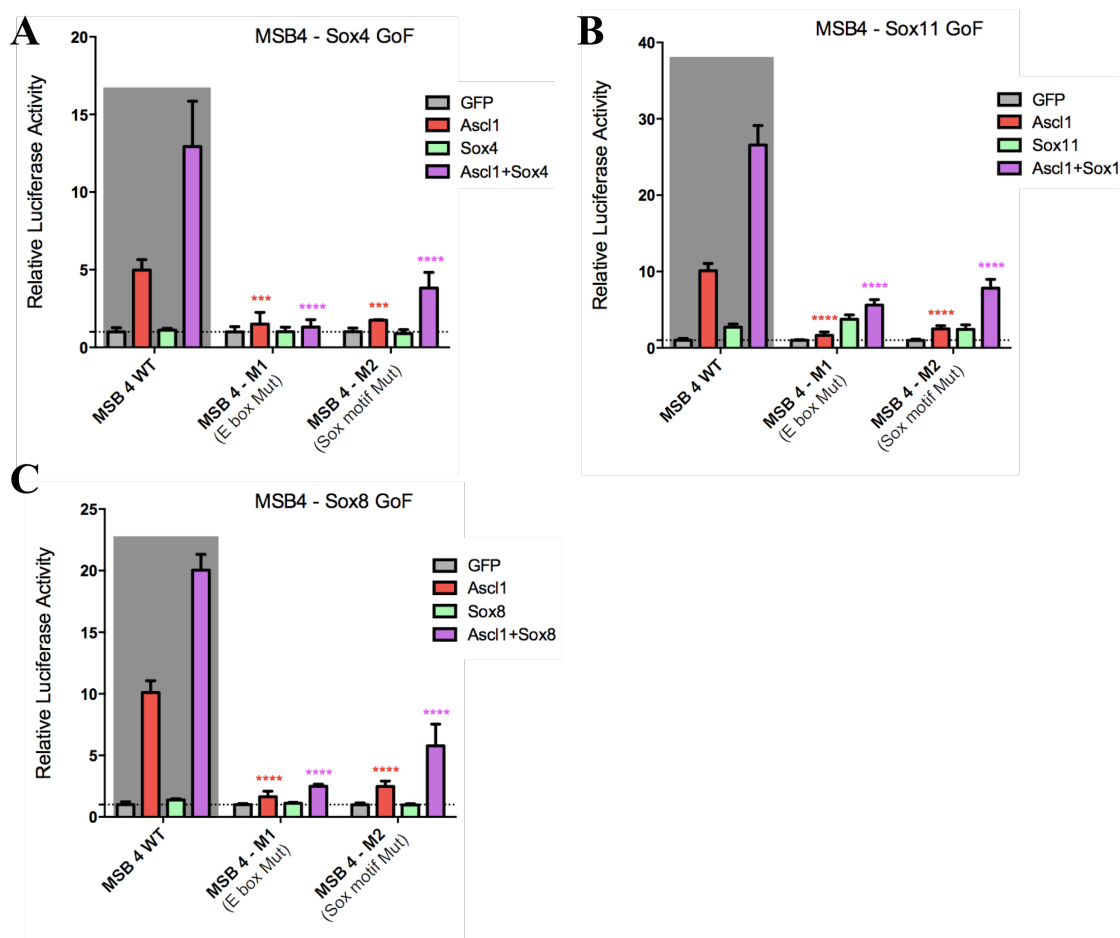


Figure 4.5. Mutations of E-box and Sox motifs affect the synergy of Ascl1 with SoxC factors, or Sox8 in the activation of neuronal enhancer MSB4.

Luciferase activity of mutants and wt enhancer MSB4 was measured and compared after overexpression of Ascl1, SoxC or Sox8 either alone or simultaneously (as shown in each legend near the graphs). Luciferase activity is expressed as relative compared to the activity of the same enhancer after overexpression of the GFP vector as control. Luciferase activity of the wt enhancer is shown on grey background in A, B, and C for comparison with the mutant elements. The neuronal enhancer MSB4 is synergistically activated by simultaneous overexpression of Ascl1 and Sox4/Sox11 or Sox8 (A, B, and C, wt element). Mutation of the E-box abolished activation of the enhancer by exogenous Ascl1 and also synergistic activation by Ascl1 and Sox4 (A), Ascl1 and Sox11 (B), and Ascl1 and Sox8 (C) (element M1 vs wt). Mutation of the Sox motif also abolished activation of the enhancer by exogenous Ascl1 as previously demonstrated (element M2 vs wt element in A, B, and C) and also the synergistic activation of the enhancer by Ascl1 and SoxC or Sox8 (element M2 vs wt in A, B, and C). Mutation of the Sox motif did not change the luciferase activity of mutant element M2 after overexpression of the three Sox factors, Sox4 in A, Sox11 in B, Sox8 in C. The data are presented as the mean \pm SD of triplicate assays.

4.5.2 Regulation of MSB22-short mutant enhancers by Ascl1 and Sox factors

The neuronal enhancer MSB22-short is also activated by overexpression of Ascl1 in NS5 cells. Also in this case as for the other neuronal enhancers analysed, MSB22-short is synergistically activated by simultaneous overexpression of Ascl1 and Sox4/Sox11. In contrast, Sox8 inhibits Ascl1-induced activation of the enhancer when these two TFs are overexpressed together. The Sox factors do not activate the enhancer when they are overexpressed alone in NS5 cells (Figures 4.7, wt elements). Mutation of the E-box reduced activation of the enhancer by exogenous Ascl1 and also the synergy between Ascl1 and Sox4/Sox11 (Figures 4.7, mutant element M1). Mutation of Sox motif 1 did not change the regulation of the enhancer by the Sox factors and the synergy between Ascl1 and Sox4/Sox11 was not compromised (Figures 4.7, mutant element M2). Also for enhancer MSB22-short, the results of the experiments of mutagenesis cannot suggest if the Sox factors bind to their motif and regulate the enhancer. However, the synergy between Ascl1 and Sox4/Sox11 is not compromised when the Sox motif is mutated. In this case, also Ascl1-induced activation of the enhancer is not lost when the Sox motif is mutated. These results might suggest that Ascl1 and Sox4/Sox11 synergise as proteins in a same transactivation complex without binding of Sox4/Sox11 to the identified Sox motif.

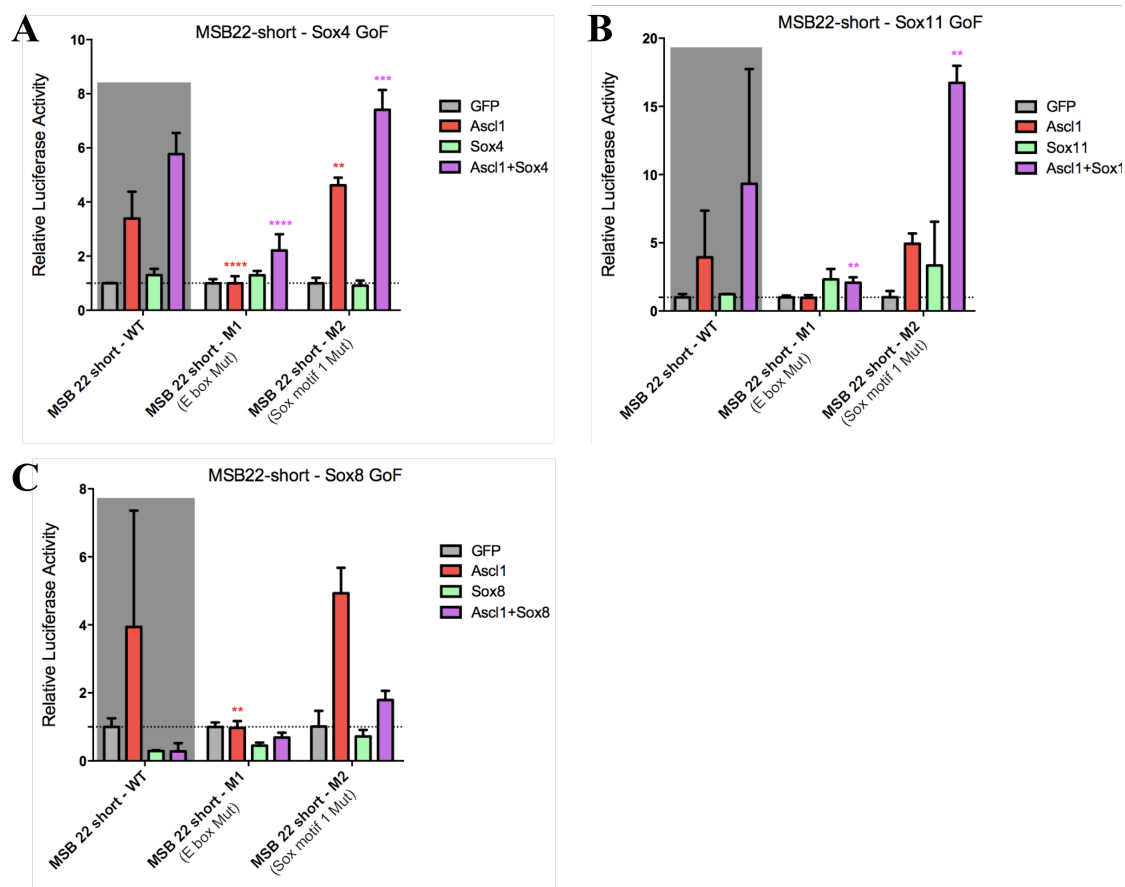


Figure 4.6. Regulation of MSB22-short mutant enhancers by Ascl1 and Sox factors.

Luciferase activity of mutants and wt enhancer MSB22-short was measured and compared after overexpression of Ascl1, Sox4 or Sox8 either alone or simultaneously (as shown in each legend near the graphs). Luciferase activity of the wt enhancer is shown on grey background in A, B, and C. Neuronal enhancer MSB22-short is synergistically activated by simultaneous overexpression of Ascl1 and Sox4/Sox11 (A and B) while Sox8 inhibits Ascl1-induced activation when these two TF are overexpressed together (C). Mutation of the E-box reduced the activation of the enhancer by exogenous Ascl1 (A and C, mutant element M1 vs wt) and also synergistic activation by Ascl1 and Sox4/Sox11 (A and B, mutant element M1 vs wt). Mutation of the Sox motif did not change the regulation by the Sox factors (A, B, and C, mutant element M2 vs wt) nor compromised the synergistic activation between Ascl1-Sox4/Sox11 (A and B, mutant element M2 vs wt). The data are presented as the mean \pm SD of triplicate assays.

4.6 Discussion

4.6.1 Identification of E-box and Sox motifs in the neural enhancers

In the previous chapter I showed that genomic elements identified through a ChIP-seq approach acted as enhancers regulated by Ascl1 and Sox2. In this chapter I focused on the identification of TF binding motifs in the enhancer sequences and their role in the regulation. One of the features of the enhancers is the presence of a dense clustering of multiple motifs where TFs bind and regulate the enhancer activity (Dickel et al., 2013, Spitz and Furlong, 2012). Using the TESS software and searching for E-box and several Sox motifs, I identified BMs for Ascl1 and Sox factors. Mostly, these were located in the regions of overlapping ChIP-seq peaks of Ascl1 and Sox2 (Figure 4.1). This finding further validates ChIP-seq data and supports the possibility that Ascl1 and Sox factors bind in close proximity to a core region of the regulatory sequences. The results described in this chapter show that mutation of the BMs affect enhancer activity for some of the regulatory regions characterised. These results suggest mechanisms of direct DNA binding of the TFs involved in the regulation of the neural enhancers identified in this study.

4.6.2 Mutagenesis of the BMs suggests regulation of the neural enhancers by endogenous bHLH factors and Sox factors

Some experiments here presented aimed to test and compare the basal transcriptional activity of mutant enhancers and their wt sequences in neural stem cells. The purpose was not only to understand if an intact motif was required for enhancer activity, but also to infer the role of the endogenous TFs, in particular endogenous Ascl1 and Sox2, in the regulation of the enhancers in neural stem cells. Luciferase assays results showing a

reduced basal activity of mutant enhancers carrying disrupted E-boxes and Sox motifs compared with the activity of the wt sequences suggest that endogenous bHLH factors and Sox factors, possibly Ascl1 and Sox2, account for the enhancer activity observed in neural stem cells. These results also suggest that bHLH factors and Sox factors regulate basal enhancer activity through direct DNA binding to the identified motifs.

The transcriptional basal activity of the mutant and wt elements to infer the role of endogenous bHLH and Sox factors in the regulation of the identified neural enhancers is particularly valuable in the absence of knock down experiments. Indeed gain of function of Ascl1 and Sox2 in NS5 cells is the only successful approach undertaken in this study to prove the effect of the two TFs in the regulation of the enhancers. Experiments of knock down of these TFs have been tried unsuccessfully in this project (data not shown).

The results described suggest that endogenous bHLH and Sox factors, possibly Ascl1 and Sox2, regulate the basal activity of enhancers MSB18 and MSB22 in neural stem cells. Mutations of specific motifs, but not all of them, were functionally deleterious and caused reduction of the basal enhancer activity. These results also suggest that endogenous bHLH and Sox factors could regulate the basal activity of enhancers MSB18 and MSB22 through direct DNA binding to the identified motifs. MSB18 and MSB22 were also among the enhancers with the highest basal activity, so the result of the mutagenesis study here carried out is consistent with the activity displayed by these enhancers in neural stem cells. Mutations of the E-box and Sox motifs in enhancer MSB4 did not affect its transcriptional basal activity. The result is consistent with the lack of activity of this regulatory element in neural stem cells. MSB4 has been classified as a neuronal enhancer since it is only activated by exogenous Ascl1 in NS5 cells. For this enhancer, combined mutation of E-box and Sox motif in the same mutant element

caused an increase of the basal activity compared with the wt element, suggesting that TFs binding to these motifs may actually repress the activity of this enhancer in neural stem cells. Finally, mutagenesis analysis of disrupted binding motifs in enhancers MSB24 and MSB11 suggest that the BMs identified and mutated either did not account or gave only a minor contribution to the basal transcriptional activity of these elements in neural stem cells.

4.6.3 Mutagenesis of the BMs suggests different mechanisms of regulation of the neural enhancers by exogenous Ascl1 and Sox2

Experiments where the effect of the overexpression of Ascl1 and Sox2 were tested on mutant enhancers compared with the wt element aimed to gain insight into the possible direct binding of the TFs to their DNA binding motifs in the enhancer sequence, and to clarify any interaction between Ascl1 and Sox2 in the regulation of the enhancers.

Loss of the E-box strongly affected exogenous Ascl1 activation of the neuronal enhancers MSB4, MSB18 and MSB22, and in a modest measure MSB11. These results give evidence that Ascl1 might activate these regulatory regions through direct DNA binding to the identified E-box.

Overall, mutation of the Sox motifs failed to show requirement of these sites for the regulation of the enhancers by Sox2 and its direct binding to the DNA, with the exception of the NSC active enhancer MSB24-short. It is possible that Sox2 regulates these elements through alternative mechanisms. It might bind to non-canonical motifs and/or regulate without direct binding to the identified motifs. These mechanisms need to be addressed. However, results of the mutagenesis of the Sox motifs and its effect on the transcriptional activity are consistent for the neuronal enhancers MSB4, MSB18, and MSB22. These enhancers did not show activation by exogenous Sox2. It is

therefore consistent that mutation of the Sox motif did not affect the regulation by this TF. Similar results were obtained by Kamachi and colleagues in their study on the activation of the DC5 enhancer of the δ -crystallin gene by Pax6 and Sox2 (Kamachi et al., 2001). In that work, no differences were observed in the values of luciferase activity of the wt and mutant DC5 enhancer carrying mutation of the Sox motif after overexpression of Sox2 in cultured liver cells. Sox2 could not activate the enhancer when overexpressed alone. However, strong activation was seen in synergy with Pax6 and cooperative binding was demonstrated for the two TFs on the DC5 enhancer. In conclusion, binding of Sox2 cannot be excluded from our results and further experiments are needed to understand the mechanism of regulation of the neuronal enhancers by this TF. Nevertheless, mutation of the Sox motif affected activation of the neuronal enhancer MSB4 by exogenous Ascl1 suggesting the requirement and dependence of a Sox motif for Ascl1 induction. This intriguing result might suggest that a protein-protein interaction between Ascl1 and a Sox factor binding to the identified Sox motif is required for Ascl1-induced activation of this neuronal enhancer. The result might also suggest a possible cooperative DNA binding to their motifs of the two proteins involved in the regulation of this enhancer. The mechanism emerging needs to be further addressed and investigated, also to conclude if it occurs as a general mechanism of regulation on other neuronal enhancers.

4.6.4 Recruitment of different TFs to different BMs in enhancer MSB18

The analysis of the five MSB18 mutant enhancers allowed discrimination between the five binding motifs identified, with recruitment of different bHLH and Sox factors involved in the regulation of this enhancer. Indeed, only mutation of E-box 1 reduced

Ascl1 activation by 4-fold, suggesting binding of the TF to this motif. While mutations of the E-box 2, E-box 3 and Sox motif 2 strongly increased activation by exogenous Ascl1 compared with the wt element, suggesting that other TFs binding these motifs might counteract and repress Ascl1 induction of the enhancer. The results give evidence of Ascl1 binding to the E-box 1, while other bHLH TFs bind to E-box 2 and E-box 3 competing with Ascl1 activation of the neuronal enhancer. When point mutation in these other E-boxes prevented these TFs from binding the DNA, Ascl1 transcriptional activation of the enhancer increased, as showed for mutant elements M2 and M3 compared with the wt element (Figure 4.3C and D). These results are consistent with results of the basal activity of the mutant elements of enhancer MSB18 since also in that case mutation of E-box 1 did not affect the basal transcriptional activity of the element. This suggests that this motif and the TF there recruited does not account for the basal activity of MSB18 in neural stem cells. While TFs binding E-box 2 and 3, and Sox motifs 2 regulate the basal activity of enhancer MSB 18. In conclusion, the results suggest that other bHLH factors rather than Ascl1 are responsible for the basal activity of enhancer MSB 18 in neural stem cells, together with a Sox factor binding to Sox motif 2. The bHLH factor Olig2 might be one of the candidates to regulation of this enhancer. Olig2 and Ascl1 bind to the same E-box. Olig2 ChIP-seq data in NS5 cells from Ben Martynoga indicate recruitment of this TF to the neural enhancers identified in this study (Martynoga, unpublished data). Moreover, luciferase assays performed in Olig2 conditional knock out cell line, derived from primary culture, exhibited a strong loss of basal activity of enhancer MSB18 when Olig2 was deleted compared with the wt cells (data not shown).

4.6.5 SoxC and Sox8 factors in the regulation of the neuronal enhancers and synergy and counteraction with Ascl1

Mutagenesis of the neuronal enhancer MSB4 showed the requirement of the Sox motif for Ascl1-induced activation of this element. The fact that overexpression of Sox2 did not activate the neuronal enhancers and counteracted Ascl1 induction when the two TFs were overexpressed together raises the possibility that other Sox factors, rather than Sox2, might bind to the same Sox motif and synergise with Ascl1 in the activation of these enhancers. This hypothesis is supported by the fact that Sox factors tend to bind to a common motif (Lefebvre et al., 2007).

RNA-seq data generated in our lab showed that Sox4, Sox11 and Sox8 were transcribed in NS5 cells, with Sox4 being the highest transcribed, and microarray data following Ascl1 GoF showed induction of Sox4, Sox11 and Sox8, and repression of Sox2 (Ben Martynoga, unpublished data). This prompted us to investigate a possible interaction between Ascl1 and Sox4, Sox11 and Sox8 in the regulation of the neuronal enhancers. Moreover, the literature reported that SoxC members, in particular Sox4 and Sox11, have a critical role in the specification of neuronal traits downstream of bHLH factors (Bergsland et al., 2006) and target the promoters of genes that are induced upon neuronal differentiation of adult NSCs, such as the doublecortin (DCX) promoter (Mu et al., 2012).

The results described in this chapter showed that Ascl1 strongly synergised with Sox4 and Sox11 in the activation of the neuronal enhancers MSB4, MSB18, MSB22 and MSB22-short when the two TFs were overexpressed together. Ascl1 synergised also with Sox8 in the activation of enhancers MSB4 and MSB18. However, Sox8 inhibited the activity of enhancer MSB22, and its shorter construct MSB22-short, by

counteracting Ascl1 induction of this enhancer when the two TFs were overexpressed simultaneously. Therefore, Sox8 was able to act as both an activator and a repressor on different neuronal enhancers. In particular, Sox8 displayed a similar pattern as Sox2 in counteracting Ascl1 on enhancer MSB22. Indeed, the result for MSB22 was consistent with data presented in the literature where Sox8 was showed to act as negative regulator of myogenin expression, repressing the activity of the myogenin promoter and abolishing bHLH MyoD-induced activation of the same element when the two TFs where overexpressed together in skeletal muscle differentiating cells (Schmidt et al., 2003).

Despite the strong synergy with Ascl1, or the counteraction in the case of Sox8, Sox4/Sox11 and Sox8 did not show regulation of the enhancers when they were overexpressed alone.

4.6.6 Two different mechanisms of synergistic activation of the neuronal enhancers by Ascl1 and the SoxC factors

Mutagenesis experiments described at the end of this chapter aimed to demonstrate if point mutations of the binding motifs could affect the synergy between Ascl1 and Sox4, Sox11 and Sox8 (where it was synergising with Ascl1 on MSB4 and MSB18). This would suggest direct binding of the Sox factors on the Sox motifs identified and mutated, and also that the synergy requires direct DNA binding of both TFs, Ascl1 and Sox factors, to their BMs. It might be possible that the binding of both TFs is required for the establishment of protein interactions in the synergistic activation of the neuronal enhancers. Results obtained were complex and different for each enhancer studied and suggest that more than one mechanism might be involved.

It is possible that mechanisms of interaction of Ascl1 and Sox factors are different for different elements analysed. For instance, for enhancer MSB4, the synergy between Ascl1 and Sox4 or Sox11 was compromised either with mutation of the E-box or with mutation of the Sox motif. In this second case, it is not possible to distinguish if the synergy is lost as a consequence of the effect of the mutation of the Sox motif on the activity of the exogenous Ascl1, or because the mutation of the Sox motif affects binding of the Sox factors. Since binding of the Sox factors cannot be excluded at this stage, these results can be compatible with a mechanism of regulation where both Ascl1 and the SoxC factors bind to their DNA motifs, probably in a cooperative way, and synergise in the activation of the enhancer that they co-bind and co-regulate. This interaction between Ascl1 and Sox4/Sox11 suggests a feedforward loop network in the regulation of the neuronal enhancer MSB4. In a feedforward loop motif, a regulator activates a second regulator and they both bind a common regulatory target that they activate together (Mangan and Alon, 2003, Boyer et al., 2005). Indeed, it is known that Ascl1 activates SoxC factors (Bergsland et al., 2006) (and Martynoga, unpublished data). It seems possible then that SoxC factors feed forward to help the same Ascl1 to activate the neuronal enhancer MSB4 (Figure 6.1B). The mechanism of regulation could be different for enhancer MSB22. In this case, the synergy between Ascl1 and SoxC factors was compromised only with mutation of the E-box but not with mutation of the Sox motif. Therefore, it is possible that Ascl1 and Sox4 or Sox11 synergise in the activation of this enhancer as proteins of a same transactivation complex without binding of the SoxC factor to the DNA (Figure 6.1C). Binding of the SoxC factors to the identified Sox motifs need to be demonstrated or excluded to further support these two models.

Mutagenesis analysis could not suggest direct binding of the Sox factors tested in this study. However, all Sox factors tested show low or no activation of the enhancers studied when they are overexpressed alone in NS5 cells. It is known that Sox factors tend to activate a regulatory region in interaction with a partner factor rather than alone (Yuan et al., 1995, Kamachi et al., 2001, Kondoh and Kamachi, 2010). If binding of the Sox factors to their motifs does not correlate with activation of the enhancers, it is possible that mutation of the same motifs disrupting the binding will not show effect on the luciferase activity of the Sox-motif mutant enhancers compared with the wt elements (luciferase activity of both wt and Sox-motif mutant enhancers is low when Sox factors are overexpressed alone and it is not affected from the mutation).

Finally, mutation of some Sox motifs, such as Sox motif 2 (in mutant element M5) in enhancer MSB18, increased exogenous *Ascl1* activation of the enhancer, suggesting that a Sox factor, not yet identified in this study, could repress *Ascl1* induction of the enhancer. This result further demonstrates that many Sox factors could be involved in the regulation of the neural enhancers, some activating and some other repressing the transcriptional activity of the regulatory elements. It also demonstrates the existence of both activating and repressing modules within the same enhancer, and a complex interplay of several TFs bound and involved in their regulation.

In conclusion, this study has identified a complex interplay of *Ascl1* with different Sox factors in the regulation of neural enhancers. Results show that *Ascl1* and Sox2 counteract each other, whereas *Ascl1* and SoxC factors synergistically activate these elements. However, the mechanisms of these antagonistic and synergistic interactions need to be further investigated and the requirement of a Sox motif for *Ascl1*-induced activation of the enhancer MSB4 needs to be explained. Moreover, mechanisms of regulation of the neural enhancer by Sox factors need to be further dissected. In

particular, direct DNA binding of the Sox factors to their motifs needs to be demonstrated or excluded. EMSA experiments can prove or exclude direct interactions of the TFs with the DNA. EMSA experiments are described in the next chapter.

5. *In vitro* binding of Ascl1 to the neuronal enhancer MSB4

5.1 Introduction

In the previous chapter I proved that TF binding motifs in most of the enhancers characterised were involved in their regulation, suggesting a direct recruitment of Ascl1 to the E-box. However, the mechanism of regulation by Sox factors appeared more elusive. In this chapter I present EMSA (Electrophoretic Mobility Shift Assay) experiments to confirm or rule out direct binding of TFs to their DNA motifs, proposing models of mechanisms of transcriptional regulation of the identified neural enhancers. EMSA is a simple and sensitive assay to detect interactions between DNA binding proteins, in this case TFs, and nucleic acid sequences. It is based on the principle that when a protein binds a nucleic acid probe, it can retard its migration through a native polyacrylamide gel compared with the migration pattern of the free unbound probe. Therefore the formation of a protein-DNA complex can be detected on an electrophoresis gel. Under appropriate experimental conditions, both qualitative and quantitative nature of the binding can be analysed, making the technique a useful tool to characterise TFs-DNA interactions, as described since the first applications of the method (Hellman and Fried, 2007) (Fried and Crothers, 1981) (Garner and Revzin, 1981). In particular for the purposes of this thesis, EMSA experiments may demonstrate if Ascl1 and/or the Sox factors analysed in this study can bind to the enhancer regions and if they can bind simultaneously. Using enhancer constructs carrying mutations in the consensus motifs for the TFs may confirm if binding occurs specifically through the predicted sites, as a mutated site should prevent a band shift from being observed. By titrating different amounts of each TF it may be possible to propose mechanisms of

interaction between the TFs. For instance, if there is cooperative binding, where binding of one factor helps recruitment of the other, or competitive binding if the binding of one may inhibit the binding of the other. For instance, if binding of a TF helps the binding of a second TF, a lower concentration of the second TF is required to cause a band shift compared to the concentration needed when the second TF binds alone. Conversely, if binding of a TF inhibits binding of the second TF then a higher concentration of the second factor is required to cause a shift and see the band compared to the concentration required when the second TF binds alone without competing.

5.1.2 Optimisation of the experimental conditions for EMSA experiments

Despite the simplicity of the technique, successful gel shifts require the optimisation of several parameters, which influence the ability of TFs to recognise their specific DNA sequences. Binding reaction conditions, gel percentage, and gel electrophoresis conditions need to be empirically optimised and determined for each TF-probe interaction to successfully detect the formation of the complexes.

The results presented in this chapter required a long optimisation process. Initially the use of biotin-labelled DNA probes was attempted for chemiluminescent detection of the complex (using Lightshift Chemiluminescent EMSA kit from Thermo Scientific), instead of the more traditional use of radioisotope-labelled probe for radioactive detection. However, this method failed and was abandoned. We suspected this was mainly due to interference of components of the rabbit reticulocyte lysate from the TnT *in vitro* transcription/translation system (see paragraph 2.5.1.1 in Materials and Methods) with the chemiluminescent detection, which hampered the clarity and detection of the signal. Moreover, the method seems to have a much lower sensitivity

compared with the traditional radioactive detection and the use of radioisotope-labelled nucleic acid probes. Next, the source of protein and the protein synthesis methods required adjustment. While the use of rabbit reticulocyte lysate of the TnT *in vitro* system for protein expression (previously mentioned) proved to be successful as a source of native Ascl1 binding the DNA probe, I could not obtain native Sox2-TnT functionally binding TF, although the protein could be detected by Western Blot (Figure 5.1 C and D; and paragraph 2.5.1.1 for description of the method). Probably this was due to misfolding of the native protein in the cell-free TNT system. I was able to obtain results only when producing Sox2 by overexpression of Sox2 plasmid in the 293T cell line, as described in paragraph 2.5.1.2 of Materials and Methods. Finally, the detection of Sox2-DNA complex required further optimisation of binding buffers, percentage of polyacrylamide gel and electrophoresis conditions to obtain and resolve the complex on the native gel. Besides these further optimisations, results obtained for Sox2 in these experiments indicate that the binding of this factor to the MSB4 probe is very weak and/or the EMSA conditions are not fully optimised for this transcription factor (data not shown). Therefore the EMSA experiments with Sox2 have not been shown in this thesis.

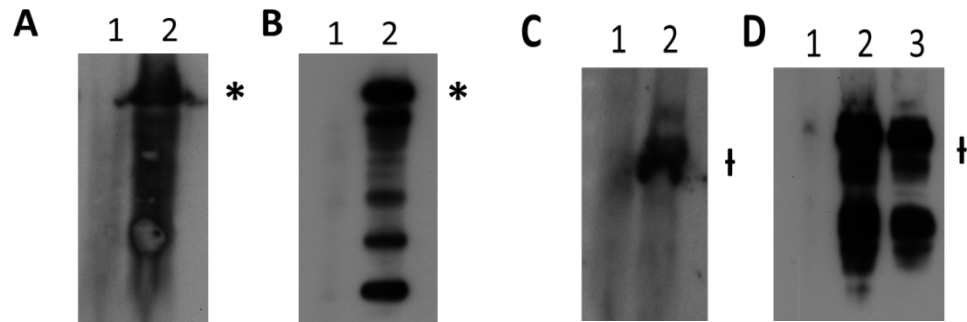


Figure 5.1 – Western blot to verify Ascl1 or Sox2 protein expression using either *in vitro* TnT system or transfection of 293T cells.

Western blot was carried out to assess expression of Ascl1 (panels A and B) or Sox2 (panels C and D). Samples were prepared using TnT system (panels A and C) or transfection of 293T cells (panels B and D). Panel A and B lanes: 1, mock; 2, Ascl1-E47. Panel C lanes: 1, mock; 2, Sox2. Panel D lanes: 1, mock; Sox2-pcDNA3.1; 3, Sox2-pCAG. * indicates position of Ascl1-E47; † indicates position of Sox2.

At the time of writing this thesis I have been able to obtain results from EMSA experiments for Ascl1 binding on the MSB 4 enhancer. I could not obtain binding of SoxC factors, in particular Sox11, on the same enhancer but also on a control probe (the Tubb3 promoter-probe used by Bergsland and colleagues) despite attempts, as conditions proved not to be optimal for the purpose (Bergsland et al., 2006). Therefore, further EMSA experiments will be needed to achieve more complete results of this chapter.

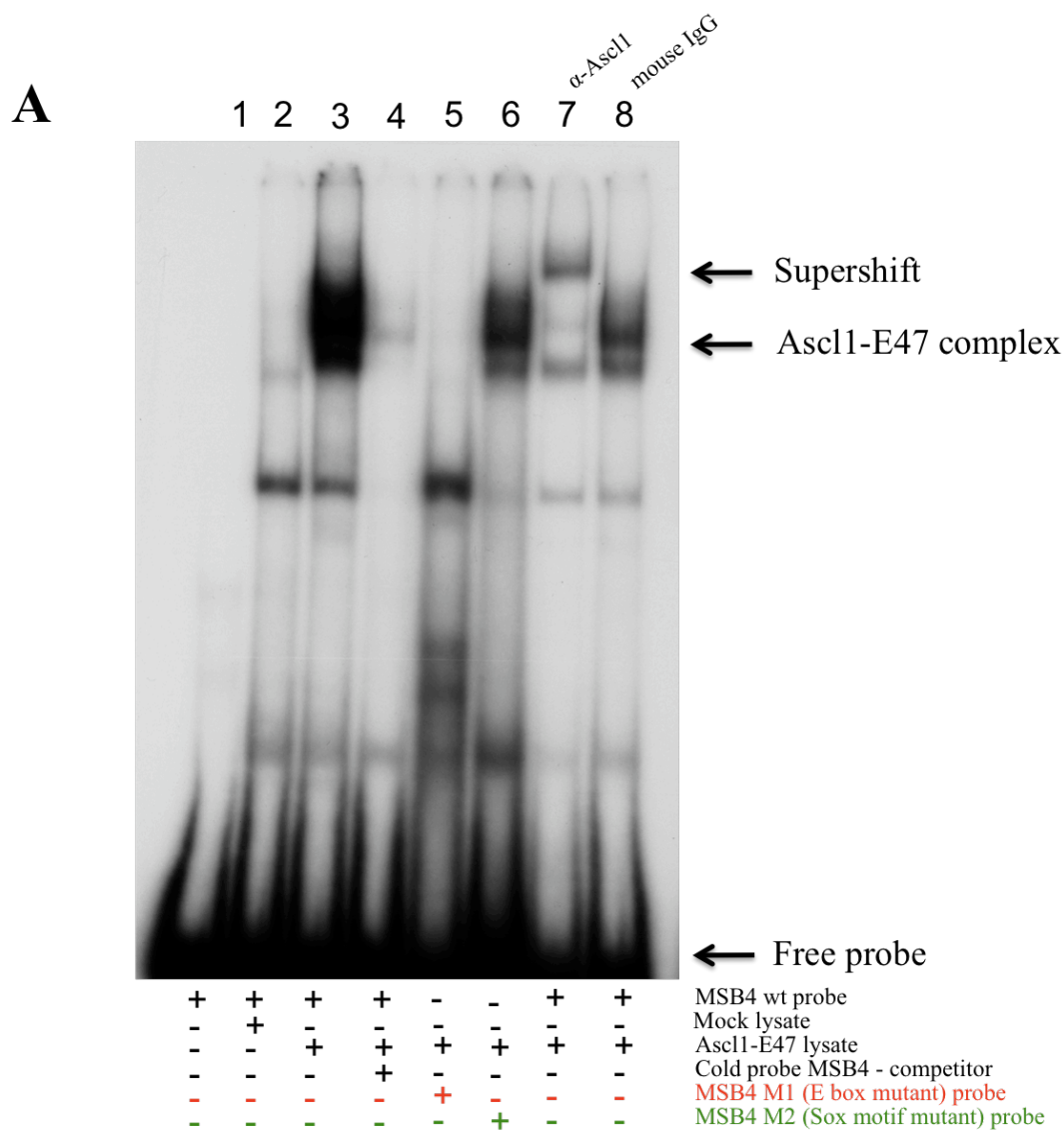
5.2 Ascl1 binds *in vitro* to the E-box in enhancer MSB4

First, I asked whether native Ascl1 was able to bind to the E-box consensus in the enhancer MSB4. I chose enhancer MSB4 for the first EMSA experiments because of its simple architecture, with only one E-box and one Sox motif present in the sequence. Furthermore, MSB4 showed an interesting feature in the regulation by Ascl1 and Sox factors, more precisely the requirement of the Sox motif for the activation of the element by exogenous Ascl1. A 63 bp ³²P-labelled DNA probe corresponding to the region of MSB4 enhancer containing both the E-box and the Sox motif, named MSB4

wt probe, was incubated with a whole cell lysate from 293T cells overexpressing Ascl1. The expression vector contained a fusion protein Ascl1 tethered to E47, which is known to bind strongly to an E-box consensus sequence (Castro et al., 2006). It is known that proneural proteins bind the DNA as heterodimers formed with the ubiquitously expressed E proteins, such as E12 and E47 in mammals (Massari and Murre, 2000, Castro et al., 2006). Therefore, the use of the fusion protein Ascl1-E47 has the advantage of strong and specific binding to the E-box in the sequence of the probe in the *in vitro* binding assay. The binding mixture was resolved on a native polyacrylamide gel. A major complex between the DNA probe and the Ascl1-293T cell lysate was detected (Figure 5.2 A, lane 3 vs 2 with Mock-293T cell lysate). To prove the identity of Ascl1 as binding protein in the complex with the DNA, a supershift assay was performed. In this control experiment, the use of an antibody against the protein of interest should recognise and bind specifically the protein. Two are the possible outcomes of the control experiment if the antibody detects and binds the protein of interest in the complex: either it further retards the migration of the complex causing a “supershift” or it competes with the DNA probe for binding to the protein and therefore weakens the intensity of the band due to reduced amount of the specific protein-DNA complex in the binding mixture. In this experiment, the addition of a mouse Ascl1 antibody led to the supershift of the complex, while the control mouse IgG antibody did not affect its mobility (Figure 5.2 A, lanes 7 and 8). Therefore, this indicates that it is specifically Ascl1-E47 to bind to the probe and to form a complex with the DNA. To prove that Ascl1-E47 binds specifically to the sequence of enhancer MSB 4, a specific competitor analysis was also carried out. The addition of a 200-fold excess of unlabelled specific probe (thus referred to as cold probe), MSB 4 cold, with the same sequence than the ^{32}P labelled one, successfully competed for Ascl1 binding, strongly

reducing the detection of the band, as showed in lane 4 (Figure 5.2 A, compare lane 4 and 3). This proves that Ascl1 tethered to E47 binds specifically to the MSB4 enhancer sequence. Next, I examined the requirement of an intact E-box and Sox motifs within the MSB4 enhancer in the formation of the Ascl1-DNA complex. Mutation of the E-box prevented Ascl1-E47 from binding when MSB4 M1 (E-box mutant probe) was added to the binding reaction mixture instead of MSB4 wt; while mutation of the Sox motif did not affect formation of the complex and detection of the band when MSB4 M2 (Sox mutant probe) was used instead of the wt probe (Figure 5.2 A, lanes 5 and 6). This proved that Ascl1 tethered to E47 needs an intact E-box for binding, and it can bind independently from the Sox motif.

Consistently, a similar experiment where the Ascl1-E47 fusion protein was expressed using the “cell-free” rabbit reticulocyte lysate TNT system for *in vitro* translation led to the same results, as showed in Figure 5.2B. In this experiment also an unrelated probe corresponding to the *Nestin* enhancer sequence, Nes probe, without known Ascl1 binding motif in the sequence, was used alternatively to the MSB4 probe to further prove sequence-specificity of the binding (Tanaka et al., 2004). Indeed, no band was observed in the absence of an Ascl1 binding motif when Nes probe was added to the binding reaction instead of MSB4 wt (Figure 5.2 B, lane 7). In conclusion, these results showed that Ascl1 binds to the MSB4 enhancer specifically through an intact E-box, and it is able to bind independently from the Sox motif.



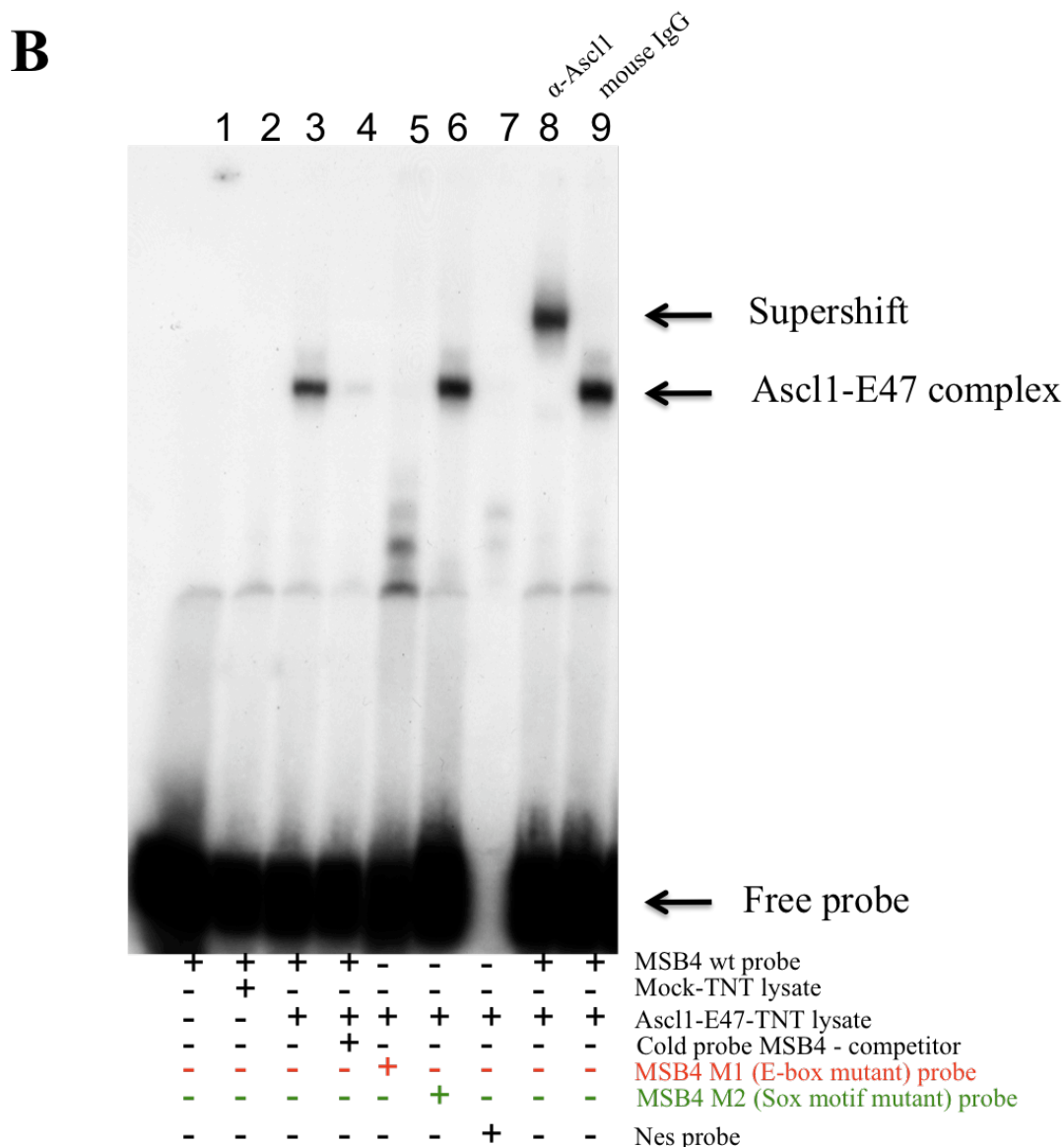


Figure 5.2. Ascl1 binds in vitro to the MSB 4 enhancer sequence.

(A) EMSA experiments were used to analyse the interaction between native Ascl1 and a 63 bp ^{32}P -labelled probe corresponding to MSB 4 wt sequence containing the E box and the Sox motif. Whole cell lysate from 293T cells transfected with a Ascl1-tethered to E47 construct was used as source of the protein. Fusion protein Ascl1-E47 was capable to form a specific complex with the wt probe as confirmed by supershift analysis with mouse anti-Ascl1 Antibody (α -Ascl1) (lanes 3, 7 and 8). 200x-fold excess of cold specific competitor probe, MSB 4 wt cold, could compete for binding with Ascl1, as proved by strong fainting of the band (lane 4). Use of a MSB 4 probe mutated in the E box, MSB 4 M1, could prevent the binding, showing requirement of an intact E box (lane 5) while the use of a MSB 4 probe mutated in the Sox motif, MSB 4 M2, did not affect the binding, proving that Ascl1 can bind independently from the Sox motif on the MSB 4 enhancer sequence (lane 6). Lane 1 shows the migration pattern of the free unbound wt probe, MSB 4 WT. Lane 2 is the control mock lysate from untransfected 293T cells. (B) The same EMSA experiment as in figure A was carried out using rabbit reticulocyte lysate TNT system as source of the fusion protein Ascl1-E47. Results were consistent in both cases. In lane 7 an unrelated probe was used, Nes probe, to confirm sequence-specificity of the binding, as no band was detected when using this probe, according to predictions.

5.3 Discussion

The results of the EMSA experiments proved that the TF Ascl1 binds directly to the MSB4 enhancer sequence. The use of mutant probes carrying aberrant E-box and Sox motif proved that binding of Ascl1 requires specifically an intact E-box and is independent from the Sox site. The results for Ascl1 are consistent with its transcriptional activation of the same enhancer, as shown in luciferase assay. Furthermore, in luciferase assay loss of activation of the mutant enhancer with disrupted E-box by Ascl1 correlates with loss of binding of the same TF to the mutated probe observed in EMSA.

In the future, it will be important to optimise the experimental condition of EMSA to determine if the Sox factors analysed in this study, in particular Sox2 and the SoxC factors Sox4 and Sox11, bind directly to the Sox motif of the same neuronal enhancer MSB4. It will be important to determine if they bind independently or simultaneously and cooperatively with Ascl1. Results from luciferase assays so far suggest that Ascl1 and Sox2 might not bind to neuronal enhancer MSB4 simultaneously since they counteract each other in the transcriptional regulation of this element. Conversely, results from luciferase assays can be compatible with cooperative binding of SoxC factors with Ascl1 to this enhancer to explain their synergistic activation of this element. In conclusion, the functionality of the Sox binding to this enhancer and the requirement of the Sox motif for the transcriptional activation of this element by exogenous Ascl1 needs to be further understood.

6. Final discussion and future work

Complex relationships between transcription factors and their enhancers and the resulting transcriptional output have profound impact in the gene regulation during development. An emerging aspect of developmental enhancers is that they undergo progressive changes that are essential for their functions and the time of their activity during development (Spitz and Furlong, 2012). A succession of regulators belonging to the same family of TFs or interplay between synergistic or antagonistic factors on the enhancers also contribute to changes in their transcriptional output and to the spatial and temporal regulation of their activity in the progression and specification of developmental processes (Kamachi et al., 2001, Tanaka et al., 2004, Bergsland et al., 2011, Aksoy et al., 2013).

The work presented in this thesis has identified and characterised nine neural enhancers regulated by a complex interplay between *Ascl1* and *Sox* factors during NSC self-renewal and neurogenesis. This work started with the identification by ChIP-seq of genomic regions co-bound by *Ascl1* and *Sox2* in neural stem cells, raising questions about the possible interactions of these factors as transcriptional regulators on common regions. An analysis of the enhancer sequences revealed that despite tending to be relatively poorly conserved, they all bind the transcriptional coactivator p300 (Figure 3.6) and are marked by the enhancer-associated epigenetic mark H3K27ac (Martynoga et al., 2013). This data support the nature of regulatory elements for these bound regions, as previously discussed (paragraph 3.6). Luciferase assays have demonstrated that 7 out of 9 regions act as enhancers *in vitro* in NS5 cell, as shown by their basal transcriptional activity. Luciferase assays have also demonstrated that overexpression of *Ascl1* and *Sox2* have an effect on the activity of all nine enhancers studied, although in

different ways for different enhancers. The different transcriptional output identified for a small group of nine enhancers points out that there is high heterogeneity in the regulation by Ascl1 and Sox2 on different enhancers. Most of the enhancers are more responsive to exogenous Ascl1 rather than Sox2, which appears to be a weaker regulator of these elements. One of the objectives of my study was the characterisation of the interaction between these two TFs as transcriptional regulators. Interestingly, my results indicate that Ascl1 and Sox2 counteract each other in the regulation of 6 out of 9 enhancers that they co-bind, and do not act synergistically in the regulation of any of the elements identified. These results are consistent with *in vivo* activities of these TFs in neurogenesis. A previous study in chick developing neural tube has demonstrated that vertebrate neurogenesis is counteracted by SoxB1 TFs (Sox1-2-3). These proteins keep neural stem cells undifferentiated by counteracting the activity of the proneural proteins. Conversely, proneural proteins can induce neurogenesis by repression of SoxB1 factors expression in neural progenitor cells (Bylund et al., 2003).

Another main objective of my work was to dissect the molecular mechanism of regulation by the TFs Ascl1 and Sox2 of the neural enhancers identified. All of the enhancers studied contained one or more E-box and Sox motifs (Figure 4.1). Mutations of these binding motifs affected enhancer activity for most of the elements analysed, suggesting a mechanism of direct TF-DNA binding in the regulation of the enhancers. The mutagenesis analysis has also demonstrated that different motifs within the same enhancer have different functions. Both activating and repressing modules have been identified within the same enhancer. Enhancers MSB18 and MSB22 are an example of this modularity, as demonstrated by mutagenesis analysis of their motifs (Figures 4.3 C and F; and paragraphs 4.6.2 and 4.6.4). This may be partly explained by the binding of other bHLH (for instance Olig2, as discussed in paragraph 4.6.4) and Sox factors

beyond those analysed in this study to the different motifs identified, for instance for enhancer MSB18. The transcriptional output seems likely to be a sum of the activity occurring at different motifs. Moreover, an analysis of the architecture of the enhancers did not reveal any reoccurring pattern in the arrangement of the E-boxes and Sox motifs within the sequences of different enhancers. Altogether, the enhancers studied in this work seem to serve as information display for binding of different TFs and their autonomous transcriptional activity, with only few requirements of interdependent motifs, as discussed later. As an information display, close TF binding motifs and their associated TFs are independently interpreted by the transcriptional machinery (Kulkarni and Arnosti, 2003). In this perspective, the neural elements characterised resemble the billboard model of enhancer architecture if compared to the more constrained arrangement of motifs of the enhanceosome model (Kulkarni and Arnosti, 2003, Arnosti and Kulkarni, 2005, Spitz and Furlong, 2012, Dickel et al., 2013). A small group of only 9 regulatory elements display a high level of heterogeneity in the enhancer architecture and possible different mechanisms of regulation for each of them. These findings are in line with the picture emerging for mammalian developmental enhancers, which also show high degree of heterogeneity in their architecture and TF binding pattern. This variability is functional to changes to their spatio-temporal activities and regulation during development (Spitz and Furlong, 2012, Dickel et al., 2013).

One of the most intriguing findings arising from this study was that mutation of the Sox motif was shown to prevent *Ascl1*-induced activation of the neuronal enhancer MSB4 suggesting the requirement of this Sox motif for exogenous *Ascl1* activity. The fact that Sox2 counteracts *Ascl1*-induced activation of this element and that different Sox factors tend to bind to the same motif prompted us to extend our analysis to other Sox factors

that could synergise with *Ascl1* in the regulation of neuronal enhancers (Harley et al., 1994, Wegner and Stolt, 2005, Lefebvre et al., 2007). Cotransfection experiments and luciferase assays presented in this thesis demonstrated that Sox4 or Sox11 strongly synergise with *Ascl1* in the activation of the neuronal enhancers MSB4, MSB18, and MSB22 when these TFs were cotransfected in NS5 cells. However, single overexpression of Sox4 or Sox11 was not sufficient to induce enhancer activity. This result is consistent with the literature reporting that Sox proteins in general need cooperation with other TFs for efficient transcriptional activation of the regulated targets (Kondoh and Kamachi, 2010). The synergy between *Ascl1*-SoxC factors presented in this study is a novel interaction between these TFs previously unreported in the regulation of neuronal enhancers in neuronal differentiation. Previous studies have reported that Sox4 and Sox11 could directly regulate and target the promoters of genes induced in neuronal differentiation, such as the *Tubb3* promoter and the Doublecortin (*DCX*) promoter (Bergsland et al., 2006, Mu et al., 2012). However, an interaction with *Ascl1* or other proneural proteins has never been considered. In particular, Bergsland and colleagues exclude such interaction, as *Ngn2* could not regulate the *Tubb3*-LacZ reporter in their transcriptional assay experiments, which identified Sox4 and Sox11 as transcriptional activators of this promoter (Bergsland et al., 2006). Mu and colleagues instead hypothesised a possible interaction with the Brn proteins since putative binding sites for these factors were identified in the *DCX* regulatory region in proximity of the Sox11 binding sites (Mu et al., 2012) and since Sox11-Brn proteins interaction has been previously reported in the regulation of the *Nestin* enhancer in the SVZ of the developing spinal cord (Tanaka et al., 2004). In support of my results, a similar synergy between the bHLH factor MyoD and Sox11 has been described in the activation of the myogenin promoter in cell culture in an *in vitro* model of myogenic differentiation

(Schmidt et al., 2003). The transcriptional assay results presented in my thesis are remarkably similar to those reported by Schmidt and colleagues in their study.

I have previously discussed that two different models of synergistic Ascl1-SoxC activation of neuronal enhancers can be proposed, according to the results of the mutagenesis experiments (paragraph 4.6.6). In some neuronal enhancers, such as MSB22 studied here, Ascl1 and Sox4 or Sox11 were still able to synergistically activate enhancer activity in the absence of a functional Sox motif, suggesting that these factors might interact as proteins in the same transactivation complex without direct binding of the SoxC factors to the DNA. Results also suggest that Ascl1 and Sox4 or Sox11 might establish a feedforward loop in the regulation of the neuronal enhancer MSB4 through direct binding of both TFs to their DNA binding motifs and synergistic activation of this neuronal enhancer (paragraph 4.6.6). Therefore, Ascl1 activates SoxC factors and these feed forward and help the same Ascl1 to activate enhancer MSB4. The diagram in Figure 6.1 (B and C) depicts both models of synergistic activation of the neuronal enhancers. In the future, it will be important in EMSA experiments to confirm or exclude direct binding of the SoxC factors to the Sox motifs in the enhancers analysed. Additionally, coimmunoprecipitation experiments will prove if Ascl1 and SoxC factors are in the same protein complex and can interact in the regulation of the enhancers where this mechanism of interaction is supported. A feedforward loop involving Sox2, Oct4 and Nanog has been described in the regulation of genes underpinning the stem state in ESCs (Boyer et al., 2005). According to the results presented in this thesis, it seems to be possible that a feedforward loop involving Ascl1 and the SoxC factors is also essential to the transcriptional regulatory network controlling neuronal differentiation, at least for one of the neuronal enhancers analysed. Future experiments will need to prove the proposed models.

The interplay between Ascl1 and the Sox factors described in this study has been dissected through overexpression of Ascl1 and Sox factors in NS5 cells. Overexpression of Ascl1 in NS5 induces neuronal differentiation and 24 hours after overexpression of this TF, NS5 cells start to acquire neuronal characteristics (Martynoga, unpublished data). In this cellular and molecular context underpinning neuronal differentiation, Ascl1 and Sox2 tend to counteract each other in the regulation of enhancers that they both bind and regulate. Conversely, Ascl1 and Sox4 or Sox11 synergise in the activation of the neuronal enhancers (the only enhancers tested at the present in this study). It is very likely that the interplay of Ascl1 and Sox factors we describe reflects the context of neuronal differentiation. It will be interesting in the future to understand if any other possible interaction is possible for Ascl1 and Sox2. For instance, it will be interesting to understand if they might ever synergise in the regulation of commonly bound enhancers active in the neural stem cell context. The GoF approach undertaken in this study excludes the characterisation of this status. Sox2 appears as a weaker regulator of the enhancers here characterised compared with Ascl1. It is possible that Sox2 is a weak activator of this enhancer because, as other Sox factors, it requires cooperative interactions with other TFs for an efficient activation of the regulated target, as previously mentioned in this thesis several other times (Kondoh and Kamachi, 2010). It is evident that Ascl1 is not the partner required for the transcriptional activation of the regulated enhancers. However, it could be speculated that Sox2 might also have architectural role and it might bind and bend the DNA to facilitate recruitment of Ascl1 or even of the other Sox factors bound at later stage of the neuronal differentiation process (Wegner, 1999, Lefebvre et al., 2007). This might explain the requirement of the Sox motif for exogenous Ascl1-induced activation of the neuronal enhancer MSB4 not fully understood at the present in this study. In this perspective, binding of

exogenously expressed Sox2 might not have evident functional consequences in the activity of the enhancers. Pre-binding of Sox2 on enhancers that will be later activated upon binding of Sox4 or Sox11 and their synergistic interaction with Ascl1 would be another possibility in the regulation of the neuronal enhancers characterised in this study. This role would be consistent with the study of Bergsland and colleagues, who proposed a succession of Sox factors, more precisely Sox2, Sox3, and Sox11 to common targets at subsequent phases of the neuronal lineage (Bergsland et al., 2011).

A better characterisation of the cellular and molecular contexts underpinning the neural stem state and the neuronal differentiating state associated to Ascl1 overexpression is needed in the future. For instance, it would be important to test if in the cellular context of Ascl1 and Sox2 overexpression in NS5 cells Sox2 blocks Ascl1-induced neurogenesis and if SoxC factors synergise with Ascl1 to increase and accelerate the production of neurons, as suggested from the transcriptional network uncovered in this study. Furthermore, performing ChIP experiments for the Sox factors in NS5 cells before or after Ascl1 overexpression could help in the molecular characterisation of the stem state or neuronal state. In this way, a rewiring of different Sox factors induced by Ascl1 could be demonstrated in the progression of the neural lineage from neural stem cells to differentiating neurons, if this occurs.

In conclusion, this study has identified a transcriptional regulatory network dependent on Ascl1 in the regulation of neuronal enhancers in differentiating neurons. In this network, Ascl1 and Sox2 counteract each other in the regulation of neuronal enhancers. Conversely, Ascl1 synergise with Sox4 or Sox11 in the activation of the neuronal enhancers. The diagram in Figure 6.1 illustrates the models of regulation that can be proposed from the results presented in this thesis.

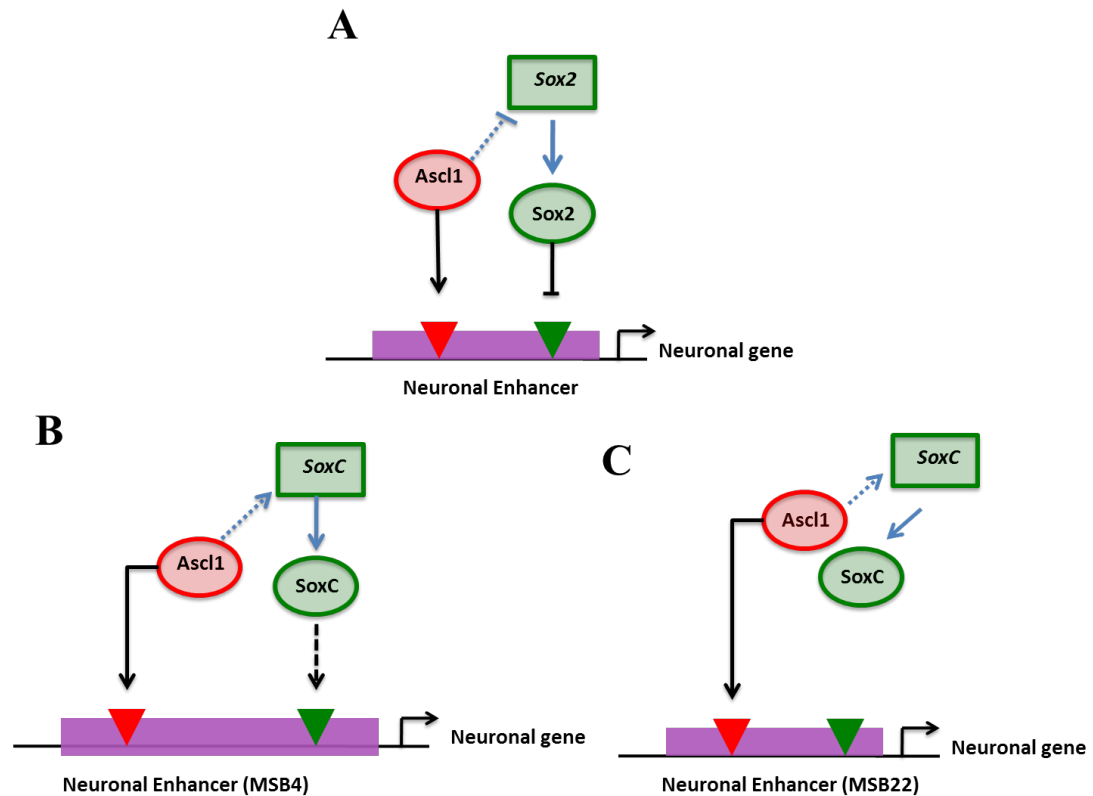


Figure 6.1 Models of interplay between *Ascl1* and Sox factors in the regulation of the neuronal enhancers.

(A) *Ascl1* and *Sox2* counteract each other in the regulation of the neuronal enhancers. (B and C) *Ascl1* and *SoxC* factors synergise in the regulation of the neuronal enhancers. (B) Feedforward regulatory loop for neuronal enhancer MSB4. *Ascl1* induces activation of the neuronal enhancer and also induces *SoxC* factors. *SoxC* factors activate the neuronal enhancer in synergy with *Ascl1* through direct DNA binding to the same regulatory sequence. The black dashed arrow indicates a direct DNA binding that still need to be tested in future experiments. (C) *Ascl1* and *SoxC* factors synergise in the activation of the neuronal enhancers, although there are no evidence of direct *SoxC* factors binding to the DNA. The ovals represent TFs; the rectangles represent genes whose names are printed in italics; the red and green triangles represent E-boxes and Sox motifs, respectively. In all three pictures the blue dashed arrows or inhibitory link have been reported in the literature (e.g. Bergslund et al., 2006) and/or have been showed in microarray experiments (data from Ben Martynoga). The black arrows showing direct binding of *Ascl1* to the E-boxes in the enhancer sequences have been directly demonstrated in EMSA experiments (MSB4 enhancer) or supported by results of mutagenesis of the E-boxes for all enhancers mutated.

References

- AKSOY, I., JAUCH, R., CHEN, J., DYLA, M., DIVAKAR, U., BOGU, G. K., TEO, R., LENG NG, C. K., HERATH, W., LILI, S., HUTCHINS, A. P., ROBSON, P., KOLATKAR, P. R. & STANTON, L. W. 2013. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J*, advance online publication.
- ALFATAH MANSOUR, A. & HANNA, J. H. 2013. Oct4 shuffles Sox partners to direct cell fate. *EMBO J*, advance online publication.
- ALVAREZ-BOLADO, G., ROSENFELD, M. G. & SWANSON, L. W. 1995. Model of forebrain regionalization based on spatiotemporal patterns of POU-III homeobox gene expression, birthdates, and morphological features. *J Comp Neurol*, 355, 237-95.
- ARLOTTA, P., MOLYNEAUX, B. J., CHEN, J., INOUE, J., KOMINAMI, R. & MACKLIS, J. D. 2005. Neuronal subtype-specific genes that control corticospinal motor neuron development in vivo. *Neuron*, 45, 207-21.
- ARMULIK, A., GENOVE, G. & BETSHOLTZ, C. 2011. Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. *Dev Cell*, 21, 193-215.
- ARNOSTI, D. N. & KULKARNI, M. M. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem*, 94, 890-8.
- AVILION, A. A., NICOLIS, S. K., PEVNY, L. H., PEREZ, L., VIVIAN, N. & LOVELL-BADGE, R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev*, 17, 126-40.
- BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27, 299-308.
- BERGSLAND, M., RAMSKOLD, D., ZAOUTER, C., KLUM, S., SANDBERG, R. & MUHR, J. 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev*, 25, 2453-64.
- BERGSLAND, M., WERME, M., MALEWICZ, M., PERLMANN, T. & MUHR, J. 2006. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes & Development*, 20, 3475-3486.
- BERNINGER, B., GUILLEMOT, F. & GOTZ, M. 2007. Directing neurotransmitter identity of neurones derived from expanded adult neural stem cells. *Eur J Neurosci*, 25, 2581-90.

- BERTRAND, N., CASTRO, D. S. & GUILLEMOT, F. 2002. Proneural Genes and the Specification of Neural Cell Types. *Nature Reviews Neuroscience*, 3, 517-530.
- BIBEL, M., RICHTER, J., LACROIX, E. & BARDE, Y. A. 2007. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat Protoc*, 2, 1034-43.
- BIBEL, M., RICHTER, J., SCHRENK, K., TUCKER, K. L., STAIGER, V., KORTE, M., GOETZ, M. & BARDE, Y. A. 2004. Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat Neurosci*, 7, 1003-9.
- BLOW, M. J., MCCULLEY, D. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAHER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., BRISTOW, J., REN, B., BLACK, B., RUBIN, E. M., VISEL, A. & PENNACCHIO, L. A. 2010. ChIP-seq identification of weakly conserved heart enhancers. *Nature genetics*, 42.
- BOYER, L. A., LEE, T. I., COLE, M. F., JOHNSTONE, S. E., LEVINE, S. S., ZUCKER, J. P., GUENTHER, M. G., KUMAR, R. M., MURRAY, H. L., JENNER, R. G., GIFFORD, D. K., MELTON, D. A., JAENISCH, R. & YOUNG, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122, 947-56.
- BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72, 248-54.
- BYLUND, M., ANDERSSON, E., NOVITCH, B. G. & MUHR, J. 2003. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci*, 6, 1162-8.
- CAHILL, M. A., ERNST, W. H., JANKNECHT, R. & NORDHEIM, A. 1994. Regulatory squelching. *FEBS Letters*, 344, 105-108.
- CAMPBELL, K. 2003. Dorsal-ventral patterning in the mammalian telencephalon. *Curr Opin Neurobiol*, 13, 50-6.
- CASAROSA, S., FODE, C. & GUILLEMOT, F. 1999. Mash1 regulates neurogenesis in the ventral telencephalon. *Development*, 126, 525-34.
- CASTRO, D. S., MARTYNOGA, B., PARRAS, C., RAMESH, V., PACARY, E., JOHNSTON, C., DRECHSEL, D., LEBEL-POTTER, M. L., GARCIA, L. G., HUNT, C., DOLLE, D., BITHELL, A., ETTWILLER, L., BUCKLEY, N. & GUILLEMOT, F. O. 2011. A novel function of the proneural factor *Ascl1* in progenitor proliferation identified by genome-wide characterization of its targets. *Genes & Development*, 25, 930-945.

- CASTRO, D. S., SKOWRONSKA-KRAWCZYK, D., ARMANT, O., DONALDSON, I. J., PARRAS, C., HUNT, C., CRITCHLEY, J. A., NGUYEN, L., GOSSLER, A., GOTTGENS, B., MATTER, J. M. & GUILLEMOT, F. 2006. Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. *Dev Cell*, 11, 831-44.
- CAVALLARO, M., MARIANI, J., LANCINI, C., LATORRE, E., CACCIA, R., GULLO, F., VALOTTA, M., DEBIASI, S., SPINARDI, L., RONCHI, A., WANKE, E., BRUNELLI, S., FAVARO, R., OTTOLENGHI, S. & NICOLIS, S. K. 2008. Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. *Development*, 135, 541-57.
- CHEN, X., XU, H., YUAN, P., FANG, F., HUSS, M., VEGA, V. B., WONG, E., ORLOV, Y. L., ZHANG, W., JIANG, J., LOH, Y. H., YEO, H. C., YEO, Z. X., NARANG, V., GOVINDARAJAN, K. R., LEONG, B., SHAHAB, A., RUAN, Y., BOURQUE, G., SUNG, W. K., CLARKE, N. D., WEI, C. L. & NG, H. H. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, 133, 1106-17.
- CHEW, J. L., LOH, Y. H., ZHANG, W., CHEN, X., TAM, W. L., YEAP, L. S., LI, P., ANG, Y. S., LIM, B., ROBSON, P. & NG, H. H. 2005. Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol*, 25, 6031-46.
- CONTI, L. & CATTANEO, E. 2010. Neural stem cell systems: physiological players or in vitro entities? *Nat Rev Neurosci*, 11, 176-87.
- CONTI, L., POLLARD, S. M., GORBA, T., REITANO, E., TOSELLI, M., BIELLA, G., SUN, Y., SANZONE, S., YING, Q. L., CATTANEO, E. & SMITH, A. 2005. Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol*, 3, e283.
- CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107, 21931-6.
- DE WIT, E. & DE LAAT, W. 2012. A decade of 3C technologies: insights into nuclear organization. *Genes Dev*, 26, 11-24.
- DICKEL, D. E., VISEL, A. & PENNACCHIO, L. A. 2013. Functional anatomy of distant-acting mammalian enhancers. *Philos Trans R Soc Lond B Biol Sci*, 368, 20120359.
- DY, P., PENZO-MENDEZ, A., WANG, H., PEDRAZA, C. E., MACKLIN, W. B. & LEFEBVRE, V. 2008. The three SoxC proteins--Sox4, Sox11 and Sox12--

exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res*, 36, 3101-17.

ENGELLEN, E., AKINCI, U., BRYNE, J. C., HOU, J., GONTAN, C., MOEN, M., SZUMSKA, D., KOCKX, C., VAN IJCKEN, W., DEKKERS, D. H., DEMMERS, J., RIJKERS, E. J., BHATTACHARYA, S., PHILIPSEN, S., PEVNY, L. H., GROSVELD, F. G., ROTTIER, R. J., LENHARD, B. & POOT, R. A. 2011. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet*, 43, 607-11.

ETTWILLER, L., PATEN, B., RAMIALISON, M., BIRNEY, E. & WITTBRODT, J. 2007. Trawler: de novo regulatory motif discovery pipeline for chromatin immunoprecipitation. *Nat Methods*, 4, 563-5.

FARAH, M. H., OLSON, J. M., SUCIC, H. B., HUME, R. I., TAPSCOTT, S. J. & TURNER, D. L. 2000. Generation of neurons by transient expression of neural bHLH proteins in mammalian cells. *Development*, 127, 693-702.

FARNHAM P. J. 2009. Insights from genomic profiling of transcription factors. *Nature Reviews Genetics*, 10.

FODE, C., MA, Q., CASAROSA, S., ANG, S. L., ANDERSON, D. J. & GUILLEMOT, F. 2000. A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes Dev*, 14, 67-80.

FRIED, M. & CROTHERS, D. M. 1981. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res*, 9, 6505-25.

GARNER, M. M. & REVZIN, A. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res*, 9, 3047-60.

GHISLETTI, S., BAROZZI, I., MIETTON, F., POLLETTI, S., DE SANTA, F., VENTURINI, E., GREGORY, L., LONIE, L., CHEW, A., WEI, C.-L., RAGOISSIS, J. & NATOLI, G. 2010. Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. *Immunity*, 32, 317-328.

GILBERT, S. F. 2006. *Developmental Biology*, Sinauer Associates.

GRAHAM, F. L., SMILEY, J., RUSSELL, W. C. & NAIRN, R. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol*, 36, 59-74.

- GRAHAM, V., KHUDYAKOV, J., ELLIS, P. & PEVNY, L. 2003. SOX2 functions to maintain neural progenitor identity. *Neuron*, 39, 749-65.
- GUBBAY, J., COLLIGNON, J., KOOPMAN, P., CAPEL, B., ECONOMOU, A., MUNSTERBERG, A., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, 346, 245-50.
- GUILLEMOT, F. & JOYNER, A. L. 1993. Dynamic expression of the murine Achaete-Scute homologue Mash-1 in the developing nervous system. *Mech Dev*, 42, 171-85.
- GUTH, S. I. & WEGNER, M. 2008. Having it both ways: Sox protein function between conservation and innovation. *Cell Mol Life Sci*, 65, 3000-18.
- HARLEY, V. R., LOVELL-BADGE, R. & GOODFELLOW, P. N. 1994. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Res*, 22, 1500-1.
- HAUBENSAK, W., ATTARDO, A., DENK, W. & HUTTNER, W. B. 2004. Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. *Proc Natl Acad Sci U S A*, 101, 3196-201.
- HEINRICH, C., BLUM, R., GASCON, S., MASSERDOTTI, G., TRIPATHI, P., SANCHEZ, R., TIEDT, S., SCHROEDER, T., GOTZ, M. & BERNINGER, B. 2010. Directing astroglia from the cerebral cortex into subtype specific functional neurons. *PLoS Biol*, 8, e1000373.
- HEINTZMAN, N. D., HON, G. C., HAWKINS, R. D., KHERADPOUR, P., STARK, A., HARP, L. F., YE, Z., LEE, L. K., STUART, R. K., CHING, C. W., CHING, K. A., ANTOSIEWICZ-BOURGET, J. E., LIU, H., ZHANG, X., GREEN, R. D., LOBANENKOV, V. V., STEWART, R., THOMSON, J. A., CRAWFORD, G. E., KELLIS, M. & REN, B. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 459, 108-12.
- HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C., CHING, K. A., WANG, W., WENG, Z., GREEN, R. D., CRAWFORD, G. E. & REN, B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, 39, 311-8.
- HELLMAN, L. M. & FRIED, M. G. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc*, 2, 1849-61.
- HENG, J. I., NGUYEN, L., CASTRO, D. S., ZIMMER, C., WILDNER, H., ARMANT, O., SKOWRONSKA-KRAWCZYK, D., BEDOGNI, F., MATTER, J. M.,

- HEVNER, R. & GUILLEMOT, F. 2008. Neurogenin 2 controls cortical neuron migration through regulation of Rnd2. *Nature*, 455, 114-8.
- JACOBS, J. L. & DINMAN, J. D. 2004. Systematic analysis of bicistronic reporter assay data. *Nucleic Acids Research*, 32, e160.
- JAUCH, R., NG, C. K., NARASIMHAN, K. & KOLATKAR, P. R. 2012. The crystal structure of the Sox4 HMG domain-DNA complex suggests a mechanism for positional interdependence in DNA recognition. *Biochem J*, 443, 39-47.
- KAGEYAMA, R., OHTSUKA, T., SHIMOJO, H. & IMAYOSHI, I. 2008. Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition. *Nat Neurosci*, 11, 1247-51.
- KAMACHI, Y., UCHIKAWA, M., TANOUCHI, A., SEKIDO, R. & KONDOH, H. 2001. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev*, 15, 1272-86.
- KAROW, M., SANCHEZ, R., SCHICHOR, C., MASSERDOTTI, G., ORTEGA, F., HEINRICH, C., GASCON, S., KHAN, M. A., LIE, D. C., DELLAVALLE, A., COSSU, G., GOLDBRUNNER, R., GOTZ, M. & BERNINGER, B. 2012. Reprogramming of pericyte-derived cells of the adult human brain into induced neuronal cells. *Cell Stem Cell*, 11, 471-6.
- KESSARIS, N., FOGARTY, M., IANNARELLI, P., GRIST, M., WEGNER, M. & RICHARDSON, W. D. 2006. Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. *Nat Neurosci*, 9, 173-9.
- KISHI, M., MIZUSEKI, K., SASAI, N., YAMAZAKI, H., SHIOTA, K., NAKANISHI, S. & SASAI, Y. 2000. Requirement of Sox2-mediated signaling for differentiation of early *Xenopus* neuroectoderm. *Development*, 127, 791-800.
- KONDOH, H. & KAMACHI, Y. 2010. SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int J Biochem Cell Biol*, 42, 391-9.
- KOYANO-NAKAGAWA, N., WETTSTEIN, D. & KINTNER, C. 1999. Activation of *Xenopus* genes required for lateral inhibition and neuronal differentiation during primary neurogenesis. *Mol Cell Neurosci*, 14, 327-39.
- KRIEGSTEIN, A. & ALVAREZ-BUYLLA, A. 2009. The glial nature of embryonic and adult neural stem cells. *Annu Rev Neurosci*, 32, 149-84.
- KULKARNI, M. M. & ARNOSTI, D. N. 2003. Information display by transcriptional enhancers. *Development*, 130, 6569-75.

- LEE, J. E., HOLLENBERG, S. M., SNIDER, L., TURNER, D. L., LIPNICK, N. & WEINTRAUB, H. 1995. Conversion of *Xenopus* ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science*, 268, 836-44.
- LEE, S., LEE, B., LEE, J. W. & LEE, S. K. 2009. Retinoid signaling and neurogenin2 function are coupled for the specification of spinal motor neurons through a chromatin modifier CBP. *Neuron*, 62, 641-54.
- LEFEBVRE, V., DUMITRIU, B., PENZO-MENDEZ, A., HAN, Y. & PALLAVI, B. 2007. Control Of Cell Fate and Differentiation by Sry-related High-mobility-group Box (Sox) Transcription Factors. *Int J Biochem Cell Biol.*, 39, 2195-2214.
- LIBER, D., DOMASCHENZ, R., HOLMQVIST, P. H., MAZZARELLA, L., GEORGIOU, A., LELEU, M., FISHER, A. G., LABOSKY, P. A. & DILLON, N. 2010. Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. *Cell Stem Cell*, 7, 114-26.
- LIN, H., MCGRATH, J., WANG, P. & LEE, T. 2007. Cellular toxicity induced by SRF-mediated transcriptional squelching. *Toxicol Sci*, 96, 83-91.
- LUI, J. H., HANSEN, D. V. & KRIEGSTEIN, A. R. 2011. Development and evolution of the human neocortex. *Cell*, 146, 18-36.
- MANGAN, S. & ALON, U. 2003. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A*, 100, 11980-5.
- MARRO, S., PANG, Z. P., YANG, N., TSAI, M. C., QU, K., CHANG, H. Y., SUDHOF, T. C. & WERNIG, M. 2011. Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell*, 9, 374-82.
- MARTYNOGA, B., DRECHSEL, D. & GUILLEMOT, F. 2012. Molecular control of neurogenesis: a view from the mammalian cerebral cortex. *Cold Spring Harb Perspect Biol*, 4.
- MARTYNOGA, B., MATEO, J. L., ZHOU, B., ANDERSEN, J., ACHIMASTOU, A., URBAN, N., VAN DEN BERG, D., GEORGOPOULOU, D., HADJUR, S., WITTBRODT, J., ETTWILLER, L., PIPER, M., GRONOSTAJSKI, R. M. & GUILLEMOT, F. 2013. Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev*, 27, 1769-86.
- MASSARI, M. E. & MURRE, C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol*, 20, 429-40.
- MASUI, S., NAKATAKE, Y., TOYOOKA, Y., SHIMOSATO, D., YAGI, R., TAKAHASHI, K., OKOCHI, H., OKUDA, A., MATOBA, R., SHAROV, A. A., KO, M. S. & NIWA, H. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol*, 9, 625-35.

- MCCONNELL, S. K. 1995a. Constructing the cerebral cortex: neurogenesis and fate determination. *Neuron*, 15, 761-8.
- MCCONNELL, S. K. 1995b. Strategies for the generation of neuronal diversity in the developing central nervous system. *J Neurosci*, 15, 6987-98.
- MIKKELSEN, T. S., KU, M., JAFFE, D. B., ISSAC, B., LIEBERMAN, E., GIANNOUKOS, G., ALVAREZ, P., BROCKMAN, W., KIM, T. K., KOCHER, R. P., LEE, W., MENDENHALL, E., O'DONOVAN, A., PRESSER, A., RUSS, C., XIE, X., MEISSNER, A., WERNIG, M., JAENISCH, R., NUSBAUM, C., LANDER, E. S. & BERNSTEIN, B. E. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448, 553-60.
- MIYAGI, S., NISHIMOTO, M., SAITO, T., NINOMIYA, M., SAWAMOTO, K., OKANO, H., MURAMATSU, M., OGURO, H., IWAMA, A. & OKUDA, A. 2006. The Sox2 regulatory region 2 functions as a neural stem cell-specific enhancer in the telencephalon. *J Biol Chem*, 281, 13374-81.
- MIYAGI, S., SAITO, T., MIZUTANI, K., MASUYAMA, N., GOTOH, Y., IWAMA, A., NAKAUCHI, H., MASUI, S., NIWA, H., NISHIMOTO, M., MURAMATSU, M. & OKUDA, A. 2004. The Sox-2 regulatory regions display their activities in two distinct types of multipotent stem cells. *Mol Cell Biol*, 24, 4207-20.
- MIYATA, T., KAWAGUCHI, A., SAITO, K., KAWANO, M., MUTO, T. & OGAWA, M. 2004. Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. *Development*, 131, 3133-45.
- MIZUGUCHI, R., SUGIMORI, M., TAKEBAYASHI, H., KOSAKO, H., NAGAO, M., YOSHIDA, S., NABESHIMA, Y., SHIMAMURA, K. & NAKAFUKU, M. 2001. Combinatorial roles of olig2 and neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. *Neuron*, 31, 757-71.
- MOLYNEAUX, B. J., ARLOTTA, P., HIRATA, T., HIBI, M. & MACKLIS, J. D. 2005. Fezl is required for the birth and specification of corticospinal motor neurons. *Neuron*, 47, 817-31.
- MU, L., BERTI, L., MASSERDOTTI, G., COVIC, M., MICHAELIDIS, T. M., DOBERAUER, K., MERZ, K., REHFELD, F., HASLINGER, A., WEGNER, M., SOCK, E., LEFEBVRE, V., COUILLARD-DESPRES, S., AIGNER, L., BERNINGER, B. & LIE, D. C. 2012. SoxC Transcription Factors Are Required for Neuronal Differentiation in Adult Hippocampal Neurogenesis. *The Journal of Neuroscience*, 32, 3067-3080.

- MURRE, C., MCCAWE, P. S. & BALTIMORE, D. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell*, 56, 777-83.
- NAKADA, Y., HUNSAKER, T. L., HENKE, R. M. & JOHNSON, J. E. 2004. Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification. *Development*, 131, 1319-30.
- NATESAN, S., RIVERA, V. M., MOLINARI, E. & GILMAN, M. 1997. Transcriptional squelching re-examined. *Nature*, 390, 349-50.
- NAYLOR, L. H. 1999. Reporter gene technology: the future looks bright. *Biochem Pharmacol*, 58, 749-57.
- NIETO, M., SCHUURMANS, C., BRITZ, O. & GUILLEMOT, F. 2001. Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. *Neuron*, 29, 401-13.
- NIWA, H., MIYAZAKI, J. & SMITH, A. G. 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet*, 24, 372-6.
- NOCTOR, S. C., MARTINEZ-CERDENO, V., IVIC, L. & KRIEGSTEIN, A. R. 2004. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci*, 7, 136-44.
- OKUMURA-NAKANISHI, S., SAITO, M., NIWA, H. & ISHIKAWA, F. 2005. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J Biol Chem*, 280, 5307-17.
- ORPHANIDES, G., LAGRANGE, T. & REINBERG, D. 1996. The general transcription factors of RNA polymerase II. *Genes Dev*, 10, 2657-83.
- PACARY, E., HENG, J., AZZARELLI, R., RIOU, P., CASTRO, D., LEBEL-POTTER, M., PARRAS, C., BELL, D. M., RIDLEY, A. J., PARSONS, M. & GUILLEMOT, F. 2011. Proneural transcription factors regulate different steps of cortical neuron migration through Rnd-mediated inhibition of RhoA signaling. *Neuron*, 69, 1069-84.
- PANNE, D. 2008. The enhanceosome. *Curr Opin Struct Biol*, 18, 236-42.
- PANNE, D., MANIATIS, T. & HARRISON, S. C. 2007. An atomic model of the interferon-beta enhanceosome. *Cell*, 129, 1111-23.

- PENNACCHIO, L. A., AHITUV, N., MOSES, A. M., PRABHAKAR, S., NOBREGA, M. A., SHOUKRY, M., MINOVITSKY, S., DUBCHAK, I., HOLT, A., LEWIS, K. D., PLAJSER-FRICK, I., AKIYAMA, J., DE VAL, S., AFZAL, V., BLACK, B. L., COURONNE, O., EISEN, M. B., VISEL, A. & RUBIN, E. M. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, 444, 499-502.
- PENZO-MENDEZ, A. I. 2010. Critical roles for SoxC transcription factors in development and cancer. *Int J Biochem Cell Biol*, 42, 425-8.
- PEVNY, L. H. & NICOLIS, S. K. 2010. Sox2 roles in neural stem cells. *Int J Biochem Cell Biol*, 42, 421-4.
- REMENYI, A., LINS, K., NISSEN, L. J., REINBOLD, R., SCHOLER, H. R. & WILMANNS, M. 2003. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev*, 17, 2048-59.
- RODDA, D. J., CHEW, J. L., LIM, L. H., LOH, Y. H., WANG, B., NG, H. H. & ROBSON, P. 2005. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem*, 280, 24731-7.
- ROSENTHAL, N. 1987. Identification of regulatory elements of cloned genes with functional assays. *Methods Enzymol*, 152, 704-20.
- SANDBERG, M., KALLSTROM, M. & MUHR, J. 2005. Sox21 promotes the progression of vertebrate neurogenesis. *Nat Neurosci*, 8, 995-1001.
- SANDELIN, A., ALKEMA, W., ENGSTROM, P., WASSERMAN, W. W. & LENHARD, B. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res*, 32, D91-4.
- SCHENBORN, E. & GROSKREUTZ, D. 1999. Reporter gene vectors and assays. *Mol Biotechnol*, 13, 29-44.
- SCHEPERS, G. E., TEASDALE, R. D. & KOOPMAN, P. 2002. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell*, 3, 167-70.
- SCHMIDT, K., GLASER, G., WERNIG, A., WEGNER, M. & ROSORIUS, O. 2003. Sox8 is a specific marker for muscle satellite cells and inhibits myogenesis. *J Biol Chem*, 278, 29769-75.
- SCHUURMANS, C., ARMANT, O., NIETO, M., STENMAN, J. M., BRITZ, O., KLENIN, N., BROWN, C., LANGEVIN, L. M., SEIBT, J., TANG, H., CUNNINGHAM, J. M., DYCK, R., WALSH, C., CAMPBELL, K., POLLEUX,

- F. & GUILLEMOT, F. 2004. Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways. *EMBO J*, 23, 2892-902.
- SESSA, A., MAO, C. A., HADJANTONAKIS, A. K., KLEIN, W. H. & BROCCOLI, V. 2008. Tbr2 directs conversion of radial glia into basal precursors and guides neuronal amplification by indirect neurogenesis in the developing neocortex. *Neuron*, 60, 56-69.
- SHIMOJO, H., OHTSUKA, T. & KAGEYAMA, R. 2008. Oscillations in notch signaling regulate maintenance of neural progenitors. *Neuron*, 58, 52-64.
- SIEGENTHALER, J. A., TREMPER-WELLS, B. A. & MILLER, M. W. 2008. Foxg1 haploinsufficiency reduces the population of cortical intermediate progenitor cells: effect of increased p21 expression. *Cereb Cortex*, 18, 1865-75.
- SIEPEL, A., BEJERANO, G., PEDERSEN, J. S., HINRICHS, A. S., HOU, M., ROSENBLUM, K., CLAWSON, H., SPIETH, J., HILLIER, L. W., RICHARDS, S., WEINSTOCK, G. M., WILSON, R. K., GIBBS, R. A., KENT, W. J., MILLER, W. & HAUSSLER, D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*, 15, 1034-50.
- SINCLAIR, A. H., BERTA, P., PALMER, M. S., HAWKINS, J. R., GRIFFITHS, B. L., SMITH, M. J., FOSTER, J. W., FRISCHAUF, A. M., LOVELL-BADGE, R. & GOODFELLOW, P. N. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, 346, 240-4.
- SOMMER, L., MA, Q. & ANDERSON, D. J. 1996. neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. *Mol Cell Neurosci*, 8, 221-41.
- SOUTHALL, T. D. & BRAND, A. H. 2009. Neural stem cell transcriptional networks highlight genes essential for nervous system development. *EMBO J*, 28, 3799-807.
- SPITZ, F. & FURLONG, E. E. M. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*, 13, 613-626.
- SUN, Y., NADAL-VICENS, M., MISONO, S., LIN, M. Z., ZUBIAGA, A., HUA, X., FAN, G. & GREENBERG, M. E. 2001. Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell*, 104, 365-76.
- SYMMONS, O. & SPITZ, F. 2013. From remote enhancers to gene regulation: charting the genome's regulatory landscapes. *Philos Trans R Soc Lond B Biol Sci*, 368, 20120358.

- TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126, 663-76.
- TANAKA, S., KAMACHI, Y., TANOUCHI, A., HAMADA, H., JING, N. & KONDOH, H. 2004. Interplay of SOX and POU Factors in Regulation of the Nestin Gene in Neural Primordial Cells. *Molecular and Cellular Biology*, 24, 8834-8846.
- TARANOVA, O. V., MAGNESS, S. T., FAGAN, B. M., WU, Y., SURZENKO, N., HUTTON, S. R. & PEVNY, L. H. 2006. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes Dev*, 20, 1187-202.
- TOKUZAWA, Y., KAIHO, E., MARUYAMA, M., TAKAHASHI, K., MITSUI, K., MAEDA, M., NIWA, H. & YAMANAKA, S. 2003. Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. *Mol Cell Biol*, 23, 2699-708.
- TOMIOKA, M., NISHIMOTO, M., MIYAGI, S., KATAYANAGI, T., FUKUI, N., NIWA, H., MURAMATSU, M. & OKUDA, A. 2002. Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res*, 30, 3202-13.
- UCHIKAWA, M., KAMACHI, Y. & KONDOH, H. 1999. Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. *Mech Dev*, 84, 103-20.
- VIERBUCHEN, T., OSTERMEIER, A., PANG, Z. P., KOKUBU, Y., SUDHOF, T. C. & WERNIG, M. 2010. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*, 463, 1035-41.
- VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJSER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009a. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457.
- VISEL, A., MINOVITSKY, S., DUBCHAK, I. & PENNACCHIO, L. A. 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res*, 35, D88-92.
- VISEL, A., RUBIN, E. M. & PENNACCHIO, L. A. 2009b. Genomic Views of Distant-acting Enhancers. *Nature*, 461.
- VO, N. & GOODMAN, R. H. 2001. CREB-binding Protein and p300 in Transcriptional Regulation. *The Journal of Biological Chemistry*, 276, 13505-13508.

- WANG, Z., BENOIT, G., LIU, J., PRASAD, S., AARNISALO, P., LIU, X., XU, H., WALKER, N. P. & PERLMANN, T. 2003. Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. *Nature*, 423, 555-60.
- WEGNER, M. 1999. From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Res*, 27, 1409-20.
- WEGNER, M. 2010. All purpose Sox: The many roles of Sox proteins in gene expression. *The International Journal of Biochemistry & Cell Biology*, 42, 381-390.
- WEGNER, M. 2011. SOX after SOX: SOXession regulates neurogenesis. *Genes Dev*, 25, 2423-8.
- WEGNER, M. & STOLT, C. C. 2005. From stem cells to neurons and glia: a Soxist's view of neural development. *Trends in Neurosciences*, 28, 583-588.
- WERNER, M. H. & BURLEY, S. K. 1997. Architectural transcription factors: proteins that remodel DNA. *Cell*, 88, 733-6.
- WERNER, M. H., HUTH, J. R., GRONENBORN, A. M. & CLORE, G. M. 1995. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, 81, 705-14.
- WINGENDER, E., DIETZE, P., KARAS, H. & KNUPPEL, R. 1996. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res*, 24, 238-41.
- WOLFFE, A. P. 1994. Architectural transcription factors. *Science*, 264, 1100-1.
- XU, J., WATTS, J. A., POPE, S. D., GADUE, P., KAMPS, M., PLATH, K., ZARET, K. S. & SMALE, S. T. 2009. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev*, 23, 2824-38.
- YUAN, H., CORBI, N., BASILICO, C. & DAILEY, L. 1995. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev*, 9, 2635-45.
- ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, 9, R137.

ZHU, Y., SUN, L., CHEN, Z., WHITAKER, J. W., WANG, T. & WANG, W. 2013.
Predicting enhancer transcription and activity from chromatin modifications.
Nucleic Acids Research.